Mitschrift zur Vorlesung

Mathematische Statistik

von Edgar Kaufmann

im Sommersemester 1998

Martin R. Elsner
$\mu_\rho \varepsilon$

16. September 1998

Inhaltsverzeichnis

I Testtheorie **4**

1 Vorbemerkungen 4

1.1 Kritische Funktionen, randomisierte Tests 6

2 Neyman-Pearson-Lemma 7

3 Tests bei monotonen Dichtequotienten 10

3.1 Einseitige Test 11

3.2 Zweiseitige Tests 13

4 Testen bei Normalverteilungen 14

4.1 Einstichprobentests bei Normalverteilungen 16

4.2 Zweistichprobentests 17

5 Optimale Konfidenzbereiche 19

II Schätztheorie **23**

6 Vorbemerkungen 23

7 Suffizienz 25

8 Erwartungstreue Schätzer und Vollständigkeit 29

8.1 Konvexe Verlustfunktionen 30

8.2 Vollständigkeit 30

9 Cramér-Rao-Ungleichung 32

III Regressionsanalyse **36**

10 Lineare Regression mit festem Design 36

10.1 Allgemeine Regression 36

10.2 Einfache lineare Regression 36

10.3 Multiple lineare Regression 37

11 Regression mit zufälligem Design 41

11.1 Lineare Regression 41

11.2 Nichtparametrische Regression 42

12 Lineare Modelle mit Normalverteilungsannahme 43

12.1 Tests bei linearen Hypothesen 46

IV Elemente der asymptotischen Statistik **50**

13 Asymptotische Normalität von Maximum-Likelihood-Schätzern 50

A Ergänzungen aus den Übungen **53**

B Literatur **55**
1 Zur Statistik 55
2 Zur Asymptotischen Statistik 55

Kapitel I

Testtheorie

1 Vorbemerkungen

Im folgenden wollen wir uns mit der *mathematischen* oder auch emphschließenden Statistik befassen (im Gegensatz zur deskriptiven Statistik). Hier sollen aufgrund von „zufälligen“ Daten Entscheidungen getroffen werden. Man versucht nun, die Wahrscheinlichkeit für das Auftreten von Fehlentscheidungen „klein“ zu halten.

Beispiel 1.1 (Punktschätzung)
In einer großen Bevölkerungsgruppe soll der Anteil der Raucher geschätzt werden. Dazu werden n Personen unabhängig voneinander befragt. Als Ergebnis erhalten wir $(x_1, \ldots, x_n) \in \{0,1\}^n$, wobei $x_i = 1$ für „i-te Person raucht“ steht. Wir fassen $(x_1, \ldots, x_n)$ als Realisierungen von i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ auf mit $X_i \sim B_{(1,p)}$, wobei p der unbekannte Anteil der Raucher ist.
Als Schätzwert liegt $\varphi(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$ nahe.
Dann ist

$$E_{B^n_{(1,p)}}(\varphi) = p, \ p \in [0,1] \tag{1.1}$$

d.h. im Mittel schätzt man den richtigen Parameter. (*Erwartungstreue*)
Frage: Wie gut ist die Schätzung?
Man kann zeigen, dass für die Varianz von φ gilt:
$Var_{B^n_{(1,p)}}(\varphi) \leq Var_{B^n_{(1,p)}}(\psi)$ für alle messbaren Funktionen $\psi : \mathbb{R} \mapsto [0,1]$, die (1.1) erfüllen, und alle $p \in [0,1]$.
Antwort: φ ist optimal, falls als Gütekriterium minimale Varianz in der Klasse der erwartungstreuen Schätzer gewählt wird.

Die Wahl des Optimalitätskriteriums hängt von der Problemstellung ab. Oft existieren auch keine optimalen Verfahren.

Beispiel 1.2 (Bereichsschätzung)
Im Beispiel 1.1 interessiert man sich auch für folgende Aussage:
Mit Wahrscheinlichkeit $1 - \alpha$ liegt der Anteil der Raucher p im Intervall
$[\varphi_1(x_1, \ldots, x_n), \varphi_2(x_1, \ldots, x_n)]$ (α klein).

Beispiel 1.3 (Testproblem)
Von einem herkömmlichen Medikament sei bekannt, dass es in 50% der Fälle heilend wirkt. Nun wird ein neues Medikament an $n = 10$ Personen getestet. Es stellen sich $k = 8$ Heilungserfolge ein. Soll das neue Medikament eingeführt werden? (Ja-Nein-Entscheidung)
Mögliche Hypothesen:
(I) Das neue Medikament ist besser als das alte Medikament.
(II) Das neue Medikament ist nicht besser.
Ziel der Untersuchung: „Statistische Absicherung von (I)“

Man nennt (II) die **Hypothese H** und (I) die Alternative oder **Gegenhypothese K**.

Mathematische Formulierung:

$(S, \mathcal{B})$	**Stichprobenraum**
Θ	**Parameterraum**
$\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$	**Familie von Wahrscheinlichkeitsmaßen**
$H \subset \Theta$	**Hypothese**
$K = \Theta \setminus H$	**Alternative**

Definition 1.4 *Ein **(nicht-randomisierter) Test** wird beschrieben durch einen **kritischen Bereich** $C \in \mathcal{B}$.*

Interpretation: Wird eine Realisation $x \in C$ beobachtet, so wird die Hypothese verworfen.

Definition 1.5 *$C \in B$ heißt **kritischer Bereich zum Niveau** $\alpha \in (0, 1)$, falls*

$$\forall \vartheta \in H : \quad Q_\vartheta(C) \leq \alpha$$

Interpretation: Ist das wahre $\vartheta \in H$, so ist die Wahrscheinlichkeit einer Fehlentscheidung kleiner oder gleich α. (**Fehler 1. Art**)

Beispiel 1.6 Sei $\alpha \in (0, 1)$, $(S, \mathcal{B}) = (\mathbb{R}^n, \mathbb{B}^n)$, $Q_\vartheta = N_{(\vartheta, \sigma^2)}$, σ fest.
$\Theta = \mathbb{R}$, $\mathcal{Q} = \{N_{(\vartheta, \sigma^2)} : \vartheta \in \mathbb{R}\}$
$H = (-\infty, \vartheta_0] \quad K = (\vartheta_0, \infty)$
$C_{n\alpha} = \{x \in \mathbb{R}^n : \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (x_i - \vartheta_0) \geq \Phi^{-1}(1 - \alpha)\}$
ist kritischer Bereich für H gegen K zum Niveau α, denn für $\vartheta \in H = (-\infty, \vartheta_0]$ gilt:
$N^n_{(\vartheta, \sigma^2)}(C_{n\alpha}) = N^n_{(\vartheta, \sigma^2)}(\{x \in \mathbb{R}^n : \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (x_i - \vartheta) \geq \Phi^{-1}(1-\alpha) + \frac{\sqrt{n}}{\sigma}(\vartheta_0 - \vartheta)\}) = 1 - \Phi(\Phi^{-1}(1 - \alpha) + \frac{\sqrt{n}}{\sigma}(\vartheta_0 - \vartheta)) \leq 1 - \Phi(\Phi^{-1}(1 - \alpha)) = \alpha$
(Für eine Verteilungsfunktion F ist $F^{-1}(q) := \inf\{t : F(t) \geq q\}$).

Definition 1.7 *$\vartheta \mapsto Q_\vartheta(C)$ heißt **Gütefunktion**.*

Ziel: Maximiere $Q_\vartheta(C)$ für $\vartheta \in K$.

1.1 Kritische Funktionen, randomisierte Tests

Definition 1.8 *Sei $\{Q_\vartheta : \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B})$ und $\Theta = H + K$.*

1. *$\varphi : (S, \mathcal{B}) \mapsto ([0,1], \mathbb{B}|_{[0,1]})$ heißt **kritische Funktion** oder **randomisierter Test**. Dabei gibt $\varphi(x)$ die Wahrscheinlichkeit an, sich beim Vorliegen der Beobachtung x für die Alternative K zu entscheiden.*

2. *φ heißt **Test zum Niveau** α **für** H **gegen** K, falls*

$$\forall \vartheta \in H : \quad E_\vartheta(\varphi) = \int \varphi \, dQ_\vartheta \leq \alpha$$

Bemerkung 1.9 (i) Ist C kritischer Bereich zum Niveau α, so ist $\varphi = 1_C$ ein Test zum Niveau α.

(ii) Ist φ ein Test zum Niveau α, so ist $C = \{x : \varphi(x) = 1\}$ kritischer Bereich zum Niveau α, denn es gilt:
$Q_\vartheta(C) = \int 1_C \, dQ_\vartheta \leq \int \varphi \, dQ_\vartheta = E_\vartheta \varphi \leq \alpha$

Definition 1.10 *Die Abbildung $\vartheta \mapsto E_\vartheta \varphi$ heißt **Gütefunktion**.*

Mögliche Entscheidungen bei Vorliegen einer Beobachtung $x \in S$:

	H wird verworfen $x \in C,\ \varphi(x) = 1$	H wird nicht verworfen $x \notin C,\ \varphi(x) = 0$
H wahr $\vartheta \in H$	Fehler 1. Art $Q_\vartheta(C),\ E_\vartheta \varphi$	richtig
K wahr $\vartheta \in K$	richtig	Fehler 2. Art $1 - Q_\vartheta(C),\ 1 - E_\vartheta \varphi$

Ist $\varphi(x) \in (0,1)$, so führt man ein unabhängiges $B_{(1,\varphi(x))}$ Zufallsexperiment aus und entscheidet sich für K bei 1 bzw. H bei 0.

Beispiel 1.11 (Fortsetzung von Beispiel1.3)
$\{B^n_{(1,p)} : p \in [0,1]\},\ p_0 = 0,5$
$H = [0, p_0],\ K = (p_0, 1]$

Der Fehler 1. Art wird als schwerwiegender betrachtet als der Fehler 2. Art.

Definition 1.12 *Seien C bzw. φ kritischer Bereich bzw. Test zum Niveau α.*

1. *φ heißt **bester Test** für H gegen $\vartheta \in K$ zum Niveau α, falls $\forall \psi$ Test z.N. α*

$$E_\vartheta \psi \leq E_\vartheta \varphi$$

2. *φ heißt* ***gleichmäßig bester Test*** *für H gegen K zum Niveau α, falls $\forall \psi$ Test z.N. α $\forall \vartheta \in K$*

$$E_\vartheta \psi \leq E_\vartheta \varphi$$

3. *C heißt* ***(gleichmäßig) bester kritischer Bereich****, falls $\varphi = 1_C$ (gleichmäßig) bester Test ist.*

Bemerkung 1.13 (i) Liefert ein Test eine Entscheidung für die Alternative, so sagt man, dass die Gültigkeit von K *statistisch gesichert* ist (zum Fehlerniveau α). Man sagt auch: Die Beobachtung x ist *signifikant* für K.
(ii) Wird H angenommen, so kann keine Aussage getroffen werden.
(iii) Will man eine Aussage *statistisch absichern*, so muss man sie als Alternative formulieren.

2 Neyman-Pearson-Lemma

Definition 2.1 *Sei Q eine Verteilung auf $(\mathbb{R},\mathbb{B})$ mit Verteilungsfunktion $F(x) = Q(-\infty, x]$. Dann heißt*

1. *$F^{-1} : (0,1) \mapsto \mathbb{R}$, $F^{-1}(q) := \inf\{t : F(t) \geq q\}$, $q \in (0,1)$,* ***Inverse Verteilungsfunktion*** *oder* ***Quantilfunktion****,*

2. *$F^{-1}(q)$* ***q-Quantil****,*

3. *$F^{-1}(1-q)$* ***q-Fraktil****.*

Wir betrachten eine einfache Hypothese H und Alternative K:
$H = \{0\}$, $K = \{1\}$, $\mathcal{Q} = \{Q_0, Q_1\}$
Q_i habe μ-Dichte f_i (i.Z. $f_i \in \frac{dQ_i}{d\mu}$), $i = 0, 1$; z.B. $\mu = Q_0 + Q_1$.

Bezeichnung 2.2 $\Phi_\alpha := \{\psi : \psi$ *Test z. N.* $\alpha\}$, $\alpha \in (0,1)$.

Ziel: Suche $\varphi \in \Phi_\alpha$ mit

$$E_1\varphi = \sup_{\psi \in \Phi_\alpha} E_1\psi$$

„Idee“: „Entscheidung für K, falls $\frac{f_1(x)}{f_0(x)}$ groß.“

Definition 2.3 *Eine kritische Funktion φ heißt* ***Neyman-Pearson-Funktion****, falls*

$$\exists c \in [0,\infty] : \ \varphi(x) = \begin{cases} 1 & f_1(x) > c f_0(x) \\ 0 & f_1(x) < c f_0(x) \end{cases}$$

Lemma 2.4 (Neyman-Pearson-Lemma) *Sei φ Neyman-Pearson-Funktion und ψ eine kritische Funktion. Dann*

1. *$E_1\psi - E_1\varphi \leq c(E_0\psi - E_0\varphi)$*

2. *Gilt* $E_0\psi \leq E_0\varphi$ *und* $E_1\psi = E_1\varphi$, *so gilt:*

$$\psi(x) = \begin{cases} 1 & \\ & falls\ f_1(x) \quad \begin{matrix} > \\ = \\ < \end{matrix} \quad cf_0(x) \qquad \mu\text{-f.ü.} \\ 0 & \end{cases}$$

3. *Sei* $\alpha \in (0,1)$ *und* $c = \inf\{y \geq 0 : Q_0\{f_1 > yf_0\} \leq \alpha\}$. *Setze*

$$\gamma = \begin{cases} \frac{\alpha - Q_0\{f_1 > cf_0\}}{Q_0\{f_1 = cf_0\}} & Q_0\{f_1 = cf_0\} > 0 \\ 0 & sonst \end{cases}$$

$$\varphi_\alpha(x) = \begin{cases} 1 & \\ \gamma & falls\ f_1(x) \quad \begin{matrix} > \\ = \\ < \end{matrix} \quad cf_0(x) \\ 0 & \end{cases}$$

Dann ist φ_α *Neyman-Pearson-Funktion und* $E_0\varphi_\alpha = \alpha$.

Beweis:

1. $(\psi - \varphi)(f_1 - cf_0) \leq 0 \quad (*)$
   Also $0 \geq \int(\psi-\varphi)(f_1-cf_0)\,d\mu = \int(\psi-\varphi)f_1\,d\mu - c\int(\psi-\varphi)f_0\,d\mu = E_1\psi - E_1\varphi - c(E_0\psi - E_0\varphi)$
2. Nach 1. gilt $c = 0$ oder $E_0\psi = E_0\varphi$. Also $\int(\psi-\varphi)(f_1-cf_0)\,d\mu = 0$. Mit (*) folgt die Behauptung.
3. $E_0\varphi_\alpha = Q_0\{f_1 > cf_0\} + \gamma Q_0\{f_1 = cf_0\} = \alpha$ (für $\gamma = 0$ siehe (+)).
   Zeige: $\gamma \in [0,1]$
   1. $Q_0\{f_1 > cf_0\} = \lim\limits_{\varepsilon\downarrow 0} \underbrace{Q_0\{f_1 > (c+\varepsilon)f_0\}}_{\leq\alpha} \leq \alpha \Rightarrow \gamma \geq 0$
   2. $Q_0\{f_1 \geq cf_0\} = \lim\limits_{\varepsilon\downarrow 0} \underbrace{Q_0\{f_1 > (c-\varepsilon)f_0\}}_{>\alpha} \geq \alpha$

   $\Rightarrow \alpha - Q_0\{f_1 > cf_0\} \leq Q_0\{f_1 = cf_0\} \Rightarrow \gamma \leq 1$
   $\gamma = 0 \Rightarrow Q_0\{f_1 = cf_0\} = 0 \Rightarrow \underbrace{Q_0\{f_1 \geq cf_0\}}_{\geq\alpha} = Q_0\{f_1 > cf_0\} \leq \alpha \quad (+)$ □

Satz 2.5 *Seien* $\alpha \in (0,1)$, $H = \{0\}$, $K = \{1\}$ *einfache Hypothese bzw. Alternative und* $\varphi = \varphi_\alpha$ *der Neyman-Pearson-Test aus Lemma 2.4. Dann gilt*

1. φ *ist bester Test zum Niveau* α, *d.h.*

$$\forall \psi \in \Phi_\alpha : \ E_1\psi \leq E_1\varphi$$

2. *Ist umgekehrt* ψ *ein bester Test zum Niveau* α, *so hat* ψ *die Form*

$$\psi = \begin{cases} 1 & \\ & f_1 \quad \begin{matrix} > \\ = \\ < \end{matrix} \quad cf_0 \qquad \mu - f.ü. \\ 0 & \end{cases}$$

Dabei ist c *das* α*-Fraktil der Verteilung von* $\frac{f_1}{f_0}$ *unter* Q_0.

Beweis:
1. folgt aus Lemma 2.4 1.
2. folgt wegen $E_1\psi = E_1\varphi$ und $E_0\psi \leq \alpha = E_0\psi$ aus Lemma 2.4 2. und 3. □

Korollar 2.6 *Sei $\varphi \in \Phi_\alpha$ ein bester Test zum Niveau α. Dann gilt:*

1. $E_1\varphi \geq \alpha$
2. $E_1\varphi > \alpha \Leftrightarrow Q_0 \neq Q_1$

Beweis: siehe Übung. □

Beispiel 2.7 (Fortsetzung von 1.3)

(i) Man teste $H = \{\mu_0\},\ K = \{\mu_1\},\ \mu_0 < \mu_1$; $Q_\vartheta = N^n_{(\mu,\sigma^2)}$
Sei φ Neyman-Pearson-Test mit $E_0\varphi = \alpha$. φ hat die Form

$$\varphi(x) = \begin{cases} 1 & > \\ f_1(x) & cf_0(x) \\ 0 & < \end{cases}, \quad x \in \mathbb{R},$$

wobei $f_j(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x_i-\mu_j)^2}{2\sigma^2}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_j)^2)$ λ^n-Dichten von $N^n_{(\mu_j,\sigma^2)}$
Weiter gilt:

$$\frac{f_1(x)}{f_0(x)} = \exp(-\tfrac{1}{2\sigma^2}\sum_{i=1}^n((x_i-\mu_1)^2-(x_i-\mu_0)^2)) = \exp(-\tfrac{1}{2\sigma^2}\sum_{i=1}^n(2x_i(\mu_0-\mu_1)+\mu_1^2-\mu_0^2)) > c$$

$$\Longleftrightarrow S_n(x) := \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (x_i-\mu_0) > \tilde{c}$$

Also hat φ die Form:

$$\varphi(x) = \begin{cases} 1 & > \\ \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (x_i-\mu_0) & \tilde{c} \\ 0 & < \end{cases}$$

mit $N^n_{(\mu_0,\sigma^2)}(\{S_n > \tilde{c}\}) = \alpha$. Also $\tilde{c} = \Phi^{-1}(1-\alpha)$ α-Fraktil von $N_{(0,1)}$.
Offenbar $N^n_{(\mu_0,\sigma^2)}\{S_n = \tilde{c}\} = 0$.

Damit ist $C_{n,\alpha} = \{x \in \mathbb{R}^n : \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (x_i-\mu_0) > \Phi^{-1}(1-\alpha)\}$ $(*)$

bester kritischer Bereich zum Niveau α für H gegen K, da $\varphi = 1_{C_{n,\alpha}}$ Neyman-Pearson-Test zum Niveau α ist.

(ii) $C_{n\alpha}$ aus $(*)$ ist auch gleichmäßig bester kritischer Bereich zum Niveau α für $H = (-\infty, \mu_0]$ gegen $K = (\mu_0, \infty)$:
Nach (i) ist $C_{n,\alpha}$ bester kritischer Bereich zum Niveau α für $H : \mu = \mu_0$ gegen $K : \mu = \mu_1$ für alle $\mu_1 > \mu_0$
Damit ist $C_{n,\alpha}$ gleichmäßig bester kritischer Bereich zum Niveau α für $H : \mu = \mu_0$ gegen

$K : \mu > \mu_0$.
Nach Beispiel 1.6 gilt $N^n_{(\mu,\sigma^2)}(C_{n,\alpha}) \leq \alpha \ \forall \mu \leq \mu_0$

Bezeichnung 2.8 *Ist φ ein Test der Form*

$$\varphi(x) = \begin{cases} 1 & > \\ & T(x) \quad c \\ 0 & < \end{cases},$$

*so heißt T **Teststatistik** und c **kritischer Wert**.*

3 Tests bei monotonen Dichtequotienten

Definition 3.1 *Seien Q_0, Q_1 Wahrscheinlichkeitsmaße auf $(S, \mathcal{B})$.*
Eine Funktion $L : (S, \mathcal{B}) \mapsto ([0, \infty], \overline{\mathbb{B}}|_{[0,\infty]})$ mit

$$Q_1(A) = \int_A L(x)\ dQ_0(x) + Q_1(A \cap \{L = \infty\}) \quad \forall A \in \mathcal{B}$$

*heißt **Dichtequotient** von Q_1 bzgl. Q_0. Schreibweise: $L \in \frac{dQ_1}{dQ_0}$*

Bemerkung 3.2 (i) Ist Q_1 durch Q_0 dominiert, dann $Q_0(\{L = \infty\}) = 0$ und L ist Q_0-Dichte von Q_1.
(ii) Ist q_i Dichte von Q_i bzgl. σ-endlichem Maß μ ($\mu \gg Q_i$), dann $L = \frac{q_1}{q_0} 1_{\{q_0>0\}} + \infty 1_{\{q_0=0\}} \in \frac{dQ_1}{dQ_0}$

Beweis: siehe Übung. □

„Rechenregel“: $\frac{a}{0} = \infty,\ a > 0, \quad 0 \cdot \infty = 0$

Definition 3.3 *Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B})$ mit $\Theta \subset \mathbb{R}$. Ferner sei $T : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ eine Statistik.*
*$\mathcal{Q}$ besitzt einen **(streng) monotonen Dichtequotienten** in T, falls gilt:*
Für jedes Paar $\vartheta_0, \vartheta_1 \in \Theta$, $\vartheta_0 < \vartheta_1$, existiert eine (streng) monotone Funktion $H_{\vartheta_0\vartheta_1} : \mathbb{R} \mapsto [0, \infty]$ derart, dass gilt

$$H_{\vartheta_0\vartheta_1} \circ T \in \frac{dQ_{\vartheta_1}}{dQ_{\vartheta_0}} \quad Q_{\vartheta_0} + Q_{\vartheta_1}\text{-f.ü.}$$

und $Q_{\vartheta_0} \neq Q_{\vartheta_1}$.

Beispiel 3.4 Folgende Familien besitzen einen monotonen Dichtequotienten:
(i) $B_{(n,p)}$-Verteilungen, $p \in (0, 1)$, $T(x) = x$.
Für $x \in \{0, \ldots, n\}$, $p_0 < p_1$:
$\frac{B_{(n,p_1)}(\{x\})}{B_{(n,p_0)}(\{x\})} = \frac{\binom{n}{x} p_1^x (1-p_1)^{n-x}}{\binom{n}{x} p_0^x (1-p_0)^{n-x}} = (\frac{p_1}{1-p_1}\frac{1-p_0}{p_0})^x (\frac{1-p_1}{1-p_0})^n = (\frac{1-p_1}{1-p_0})^n \exp(x \log(\frac{p_1}{1-p_1}\frac{1-p_0}{p_0}))$
$= (\frac{1-p_1}{1-p_0})^n \exp(\underbrace{x}_{T(x)} \underbrace{(\log \frac{p_1}{1-p_1} - \log \frac{p_0}{1-p_0})}_{>0}) =: H_{p_0,p_1}(T(x))$
(ii) $N_{(\mu,\sigma^2)}$-Verteilungen, $\mu \in \mathbb{R}$, $T(x) = x$

(iii) P_λ-Verteilung, $\lambda > 0$, $T(x) = x$

Definition 3.5 *Sei $\Theta = H + K$, φ ein Test, $\alpha \in (0,1)$*

1. *Sei $J \subset \Theta$. Der Test φ heißt **J-ähnlich**, falls gilt $\forall \vartheta \in J : E_\vartheta \varphi = \alpha$*
2. *Der Test φ heißt **unverfälscht**, falls $\forall \vartheta \in K : E_\vartheta \varphi \geq \alpha$*
   (Interpretation: Ein unverfälschter Test φ ist nicht schlechter als $\psi \equiv \alpha$)

3.1 Einseitige Test

Satz 3.6 (Optimale einseitige Tests) *$\{Q_\vartheta : \vartheta \in \Theta\}$ habe monotonen Dichtequotienten in T. Für $\alpha \in (0,1)$ betrachte das Testproblem $H = \{\vartheta \in \Theta : \vartheta \leq \vartheta_0\}$ gegen $K = \{\vartheta \in \Theta : \vartheta > \vartheta_0\}$, $\vartheta_0 \in \Theta$. Sei φ der Test*

$$\varphi(x) = \begin{cases} 1 & > \\ \gamma \quad T(x) & = \quad c \\ 0 & < \end{cases},$$

wobei c und γ bestimmt werden durch $E_{\vartheta_0}\varphi = \alpha$.

1. *φ ist ein gleichmäßig bester Test zum Niveau α für H gegen K.*
2. *φ minimiert unter allen Tests ψ mit $E_{\vartheta_0}\psi = \alpha$ ($\{\vartheta_0\}$-ähnlich) die Fehlerwahrscheinlichkeit 1. und 2. Art gleichmäßig.*
3. *Die Gütefunktion $\vartheta \mapsto \beta_\varphi(\vartheta) := E_\vartheta \varphi$ ist streng monoton steigend auf $\Theta_0 = \{\vartheta : 0 < \beta_\varphi(\vartheta) < 1\}$ und monoton auf Θ.*

Beweis:

1. Sei zunächst $\vartheta_1 > \vartheta_0$. Es gilt:
   $H_{\vartheta_0,\vartheta_1}(T(x)) > H_{\vartheta_0,\vartheta_1}(c) \Rightarrow T(x) > c$ (bzw. jeweils $<$)
   Damit hat φ die Form
   $$\varphi(x) = \begin{cases} 1 & > \\ H_{\vartheta_0,\vartheta_1}(T(x)) & \tilde{c} = H_{\vartheta_0,\vartheta_1}(c) \\ 0 & < \end{cases}$$
   Also: φ ist Neyman-Pearson-Test zum Niveau α und $E_{\vartheta_0}\varphi = \alpha$
   Mit Lemma 2.4 1. folgt, dass der Test φ optimal ist für $H = \{\vartheta_0\}$ gegen $K = \{\vartheta_1\}$.
   Da φ unabhängig von ϑ_1 ist, folgt Optimalität für $H = \{\vartheta_0\}$ gegen $K = \{\vartheta \in \Theta : \vartheta > \vartheta_0\}$
   Aus 3. folgt $E_\vartheta \varphi \leq \alpha \ \forall \vartheta \in \Theta, \ \vartheta \leq \vartheta_0$
2. Es ist zu zeigen, dass φ die Fehlerwahrscheinlichkeit 1. Art in der Klasse der $\{\vartheta_0\}$-ähnlichen Tests minimiert. Dazu betrachte das Testproblem $H = \{\vartheta_0\}$ gegen $K = \{\vartheta\}$, $\vartheta < \vartheta_0$, zum Niveau α.
   $\overline{\varphi} = 1 - \varphi$ ist Test zum Niveau $1 - \alpha$.
   Analog zu 1. zeigt man, dass $\overline{\varphi}$ Neyman-Pearson-Test ist und $E_\vartheta \overline{\varphi} = 1 - E_\vartheta \varphi$ maximiert und somit $E_\vartheta \varphi$ minimiert.
3. Seien $\vartheta_1, \vartheta_2 \in \Theta$ mit $\vartheta_1 < \vartheta_2$
   φ ist Neyman-Pearson-Test für $\{\vartheta_1\}$ gegen $\{\vartheta_2\}$ zum Niveau $\tilde{\alpha} := E_{\vartheta_1}\varphi$ (Beweis analog zu 1.).
   Aus Korollar 2.6 folgt: $\tilde{\alpha} < E_{\vartheta_2}\varphi = \beta_\varphi(\vartheta_2)$, da $Q_{\vartheta_1} \neq Q_{\vartheta_2}$ (falls $\tilde{\alpha} < 1$). □

Korollar 3.7 *Sei* $\alpha \in (0,1)$, $H = \{\vartheta \in \Theta : \vartheta \geq \vartheta_0\}$, $K = \{\vartheta \in \Theta : \vartheta < \vartheta_0\}$. *Dann erfüllt der Test*

$$\tilde{\varphi}(x) = \begin{cases} 1 & < \\ \tilde{\gamma} & T(x) = \tilde{c} \\ 0 & > \end{cases}$$

die Aussagen 3.6 1. und 2. . Dabei werden $\tilde{c}, \tilde{\gamma}$ *bestimmt aus der Gleichung*

$$\alpha = E_{\vartheta_0}\tilde{\varphi} = Q_{\vartheta_0}(\{T < \tilde{c}\}) + \tilde{\gamma} Q_{\vartheta_0}(\{T = \tilde{c}\})$$

Beweis: Man beachte, dass $\{\tilde{Q}_\vartheta : -\vartheta \in \Theta\}$, $\tilde{Q}_\vartheta = Q_{-\vartheta}$ einen monotonen Dichtequotienten in $-T$ besitzt. □

Beispiel 3.8 (Fortsetzung von Beispiel 1.3)
$S = \{0, \ldots, 10\}$, $\mathcal{B} = \mathcal{P}(S)$, $\{Q_\vartheta : \vartheta \in (0,1)\}$, $Q_\vartheta = B_{(n,\vartheta)}$
Das Testproblem lautet: $H = \{\vartheta \leq \frac{1}{2}\}$ gegen $K = \{\vartheta > \frac{1}{2}\}$. Wir wollen K zum Niveau $\alpha = 0{,}05$ statistisch absichern.
$\{Q_\vartheta : \vartheta \in (0,1)\}$ besitzt einen monotonen Dichtequotienten in $T(x) = x$. Sei F die Verteilungsfunktion von $Q_{1/2}$. Wegen $Q_{1/2}(\{x\}) = \binom{10}{x}\frac{1}{1024}$, $\binom{10}{1} = 10$, $\binom{10}{2} = 45$ gilt:

x	10	9	8	7
$F(x)$	1	$1 - \frac{1}{1024}$	$1 - \frac{11}{1024}$	$1 - \frac{56}{1024}$

Aus $\frac{56}{1024} > 0{,}05$ folgt für das α-Fraktil von F
$c_\alpha = F^{-1}(1-\alpha) = 8$
$\gamma = \frac{\alpha - Q_{1/2}(\{T > c_\alpha\})}{Q_{1/2}(\{T = c_\alpha\})} = \frac{1024}{45}(0{,}05 - \frac{11}{1024}) = \frac{1024}{45}\frac{512-110}{10\cdot 1024} = \frac{402}{450}$
Werden 9 Erfolge beobachtet, so ist die Alternative statistisch abgesichert zum Niveau $\alpha = 0{,}05$. Ist $x = 8$, so muss ein $B_{(1,\gamma)}$-Experiment durchgeführt werden.

Definition 3.9 *Sei* $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ *Familie von Wahrscheinlichkeitsmaßen auf* $(S, \mathcal{B})$.
$\mathcal{Q}$ *heißt* ***Exponentialfamilie****, falls ein* σ*-endliches Maß* μ *existiert, welches* Q_ϑ *dominiert* $\forall \vartheta \in \Theta$*, und ein* $k \in \mathbb{N}$*, Funktionen* $q_1, \ldots, q_k : \Theta \mapsto \mathbb{R}$ *und Statistiken* $T_1, \ldots, T_k : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ *existieren, so dass gilt:*

$$\forall \vartheta \in \Theta : \qquad c(\vartheta) \exp\left(\sum_{i=1}^{k} q_i(\vartheta) T_i(\cdot)\right) h(\cdot) \in \frac{dQ_\vartheta}{d\mu}$$

wobei $h : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ *und* $c(\vartheta) = (\int \exp(\sum_{i=1}^{k} q_i(\vartheta) T_i(x)) h(x)\, d\mu(x))^{-1}$.

Beispiel 3.10 (Exponentialfamilien)
(i) $\Theta = (0,1)$, $Q_\vartheta = B_{(n,\vartheta)}$, μ sei Zählmaß auf $\mathbb{N}_0$ ($\mu(\{x\}) = 1$, $x \in \mathbb{N}_0$); $x \in \{0, \ldots, n\} := S$
$B_{(n,\vartheta)}(\{x\}) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} = \underbrace{(1-\vartheta)^n}_{c(\vartheta)} \exp(\underbrace{x}_{T(x)} \underbrace{\log\tfrac{\vartheta}{1-\vartheta}}_{q(\vartheta)}) \underbrace{\binom{n}{x} 1_{\{0,\ldots,n\}}(x)}_{h(x)}$

(ii) $\Theta = \mathbb{R} \times (0,\infty)$, $Q_\vartheta = N^n_{(\mu,\sigma^2)}$, $\vartheta = (\mu,\sigma^2) \in \Theta$
Für die λ^n-Dichten gilt: (vgl. Beispiel 2.7 (i))
$\frac{dN^n_{(\mu,\sigma^2)}}{d\lambda^n}(x) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2})$
$= (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\frac{n\mu^2}{2\sigma^2}) \exp(\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2)$
Also: $(q_1(\vartheta), q_2(\vartheta)) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$, $\quad (T_1(x), T_2(x)) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$

Bemerkung 3.11 Sei $\mathcal{Q}$ eine einparametrige Exponentialfamilie mit
$c(\vartheta)\exp(q(\vartheta)T(x))h(x) \in \frac{dQ_\vartheta}{d\mu}$
Dann besitzt $\mathcal{Q}$ einen (streng) monotonen Dichtequotienten in T, falls $q : \Theta \mapsto \mathbb{R}$ (streng) monoton wachsend ist.
Beachte: $\frac{c(\vartheta_1)}{c(\vartheta_0)}\exp((q(\vartheta_1) - q(\vartheta_0))T(x)) \in \frac{dQ_{\vartheta_1}}{dQ_{\vartheta_0}}$

3.2 Zweiseitige Tests

Bemerkung 3.12 Sei $\{Q_\vartheta : \vartheta \in \Theta\}$ Familie mit monotonem Dichtequotienten, $\vartheta_0 \in \Theta$ und $Q_\vartheta \ll Q_{\vartheta_0} \forall \vartheta$
Dann gibt es keinen gleichmäßig besten Test für $H = \{\vartheta_0\}$ gegen $K = \Theta \setminus \{\vartheta_0\}$ zum Niveau $\alpha \in (0,1)$, falls $K \cap \{\vartheta \in \Theta : \vartheta > \vartheta_0\} \neq \emptyset$ und $K \cap \{\vartheta \in \Theta : \vartheta < \vartheta_0\} \neq \emptyset$.
Nehmen wir an, dass φ ein bester Test ist, so ist φ auch Test für $H = \{\vartheta_0\}$ gegen $K = \{\vartheta > \vartheta_0\}$. Also ist φ Neyman-Pearson-Test (Satz 2.5) für H gegen $\{\vartheta\}$, $\vartheta > \vartheta_0$. Wie in Satz 3.6 führt das zu

$$\varphi(x) = \begin{cases} 1 & > \\ T(x) & c_1 \\ 0 & < \end{cases} \quad \text{sowie analog (Kor. 3.7)}$$

$$\varphi(x) = \begin{cases} 1 & < \\ T(x) & c_2 \\ 0 & > \end{cases}$$

wobei Gleichheit Q_ϑ-f.s. gilt.

- Wäre $c_1 = c_2$, so folgt $T = c_1$ Q_{ϑ_0}-f.s. und somit $T = c_1$ Q_ϑ-f.s. Hieraus folgt wegen $L := H_{\vartheta_0,\vartheta} \circ T \in \frac{dQ_\vartheta}{dQ_{\vartheta_0}}$, dass L Q_{ϑ_0}-f.s. konstant ist. Das ist ein Widerspruch zu $Q_\vartheta \neq Q_{\vartheta_0}$, $\vartheta \neq \vartheta_0$
- Wäre $c_1 < c_2$, so folgt $\varphi = 1$ Q_{ϑ_0}-f.s. und $E_{\vartheta_0}\varphi = 1 > \alpha$, was ausgeschlossen ist.
- Analog $c_1 > c_2$.

Problem: (*) Teste, ob bei einer Waage die systematischen Abweichungen zwischen zwei vorgegebenen Toleranzwerten liegen.
Also $H = \{\vartheta \in \Theta : \vartheta \leq \vartheta_1 \text{ oder } \vartheta \geq \vartheta_2\}$ gegen $K = \{\vartheta \in \Theta : \vartheta \in (\vartheta_1, \vartheta_2)\}$, $\quad \Theta \subset \mathbb{R}$.

Satz 3.13 *Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ eine einparametrige, linear induzierte Exponentialfamilie mit*

$$c(\vartheta)\exp(\vartheta T(\cdot)) \in \frac{dQ_\vartheta}{d\mu}, \ \vartheta \in \Theta$$

und $TQ_\vartheta \neq \varepsilon_y \ \forall y \in \mathbb{R}$.

1. *Für das Testproblem (*) existiert ein gleichmäßig bester $\{\vartheta_1, \vartheta_2\}$-ähnlicher Test, nämlich:*

$$\varphi(x) = \begin{cases} 1 & c_1 < T(x) < c_2 \\ \gamma_i & T(x) = c_i,\ i = 1,2 \\ 0 & T(x) < c_1 \ \vee T(x) > c_2 \end{cases}$$

Dabei werden c_i und γ_i bestimmt durch $E_{\vartheta_1}\varphi = E_{\vartheta_2}\varphi = \alpha$.

2. *Der Test φ ist gleichmäßig bester Test für H gegen K zum Niveau α.*

Beweis: Siehe Witting (1985), Satz 2.69. □

Eine Anwendung finden diese Tests bei der Qualitätssicherung.

4 Testen bei Normalverteilungen

Seien $X_1, \ldots, X_n$ i.i.d. Zufallsvariablen mit $\mathcal{L}(X_i) = N_{(\mu,\sigma^2)}$.
In Beispiel 2.7 wurde das Testproblem $H = \{\mu \leq \mu_0\}$ gegen $K = \{\mu > \mu_0\}$ behandelt, wobei σ als bekannt angenommen wurde. Ein gleichmäßig bester Test zum Niveau α für dieses Testproblem ist

$$\varphi(x) = \begin{cases} 1 & \\ & \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (x_i - \mu_0) \quad \begin{matrix} > \\ \\ < \end{matrix} \quad c_\alpha \\ 0 & \end{cases}$$

wobei c_α so gewählt wurde, dass $\alpha = E_{\mu_0}\varphi = N^n_{(\mu,\sigma^2)}(\{T > c_\alpha\})$, $T(x) = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (x_i - \mu_0)$, also $c_\alpha = \Phi^{-1}(1-\alpha)$, da $TN^n_{(\mu_0,\sigma^2)} = N_{(0,1)}$.

Im Allgemeinen ist σ unbekannt. Es bietet sich an, σ durch einen Schätzer zu ersetzen, z.B.

$$s_n^2(X_1, \ldots, X_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$$

wobei $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ Schätzer für den Mittelwert der Verteilung von X_i ist, also μ.

Dies führt zu dem Test

$$\varphi(x) = \begin{cases} 1 & \\ & t(x) \quad \begin{matrix} > \\ \\ < \end{matrix} \quad \tilde{c}_\alpha \\ 0 & \end{cases}$$

mit $t(x) = \sqrt{n}\frac{\overline{x}-\mu_0}{s_n(x_1,\ldots,x_n)} = \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n (x_i-\mu_0)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2}}$, $x \in \mathbb{R}^n$.

Seien $Y_1, \ldots, Y_n$ i.i.d. mit $\mathcal{L}(Y_i) = N_{(0,1)}$.
Dann $\mathcal{L}(X_i) = \mathcal{L}(\sigma Y_i + \mu)$ und für $\mu = \mu_0$

$$\mathcal{L}(t(X_1, \ldots, X_n)) = \mathcal{L}(t(\sigma Y_1 + \mu_0, \ldots, \sigma Y_n + \mu_0)) = \mathcal{L}\left(\sqrt{n}\frac{\overline{Y}}{s_n(Y_1, \ldots, Y_n)}\right)$$

d.h. die Verteilung von t unter $N_{(\mu_0,\sigma^2)}$ hängt nicht von σ ab. Der kritische Wert c_α kann also so gewählt werden, dass φ $\{(\mu_0,\sigma^2) : \sigma > 0\}$-ähnlich ist.

Definition 4.1 *Seien $Y_0, Y_1, \ldots, Y_n$ i.i.d. mit $\mathcal{L}(Y_i) = N_{(0,1)}$. Dann heißt*

1. $\mathcal{L}(\sum_{i=1}^n Y_i^2) =: \mathcal{X}_n^2$ *(zentrale)* $\mathcal{X}_n^2$***-Verteilung mit*** n ***Freiheitsgraden***
2. $\mathcal{L}\left(\frac{Y_0}{(\frac{1}{n}\sum_{i=1}^n Y_i^2)^{1/2}}\right) =: t_n$ *(zentrale)* t_n***-Verteilung mit*** n ***Freiheitsgraden***
3. $\mathcal{L}\left(\frac{\frac{1}{s}\sum_{i=r+1}^{r+s} Y_i^2}{\frac{1}{r}\sum_{i=1}^r Y_i^2}\right) =: F_{s,r}$ *(zentrale)* $F_{s,r}$***-Verteilung mit*** $(s-r)$ ***Freiheitsgraden*** $(s \geq 1,\ r \geq 1,\ r+s \leq n)$

Satz 4.2 *Seien $X_1, \ldots, X_n$ i.i.d. mit $\mathcal{L}(X_i) = N_{(\mu,\sigma^2)}$. Dann gilt*

1. $\mathcal{L}(\sqrt{n}(\overline{X} - \mu)/\sigma) = N_{(0,1)}$
2. $\overline{X}$ *und* $\sum_{i=1}^n (X_i - \overline{X})^2$ *sind stochastisch unabhängig*
3. $\mathcal{L}(\sum_{i=1}^n (X_i - \overline{X})^2/\sigma^2) = \mathcal{X}_{n-1}^2$
4. $\mathcal{L}(\sqrt{n}(\overline{X} - \mu)/s_n(X_1, \ldots, X_n)) = t_{n-1}$

Zum Beweis zunächst

Satz 4.3 *Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit $\mathcal{L}(X_i) = N_{(\mu_i,\sigma^2)}$. Weiter seien $A \in \mathbb{R}^{n \times n}$ eine orthogonale Matrix*[1] *und $Y = AX$, wobei $X = (X_1, \ldots, X_n)^\top$, $Y = (Y_1, \ldots, Y_n)^\top$. Dann gilt:*
$Y_1, \ldots, Y_n$ sind stochastisch unabhängig mit $\mathcal{L}(Y_i) = N_{(\nu_i,\sigma^2)}$, $i = 1, \ldots, n$, wobei mit $\mu = (\mu_1, \ldots, \mu_n)^\top$, $\nu = (\nu_1, \ldots, \nu_n)^\top$ gilt: $\nu = A\mu$

Beweis: Die Abbildung $x \mapsto y := g(x) := Ax$ ist linear und es gilt $A^{-1} = A^\top$, also $g^{-1} = A^{-1}$ ebenfalls linear. Dann gilt für die Jacobi-Matrix $B(y)$ von g^{-1}:
$|B(y)| = |(\frac{\partial g_i^{-1}(y)}{\partial y_j})_{i,j}| = |A^{-1}| = 1$
Die λ^n-Dichte von $\mathcal{L}(X)$ ist
$$f(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\underbrace{\left(-\frac{\sum_{i=1}^n (x_i - \mu_i)^2}{2\sigma^2}\right)}_{\frac{1}{2\sigma^2}||x-\mu||^2}$$
Damit hat $\mathcal{L}(Y) = \mathcal{L}(AX)$ die λ^n-Dichte
$f(A^{-1}(y)) \cdot |B(y)| = f(A^{-1}(y)) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\frac{||A^{-1}y-\mu||^2}{2\sigma^2})$
Mit $||A^{-1}y-\mu|| = ||A^{-1}y-A^{-1}A\mu|| = ||A(y-A\mu)|| = ||y-A\mu|| = ||y-\nu||$ folgt die Behauptung.
□

[1] $A^{-1} = A^\top$

Damit nun

Beweis: (Satz 4.2) Sei A eine orthogonale Matrix, in deren erster Zeile $(n^{-1/2},\dots,n^{-1/2})$ steht. Dann gilt

$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i = \sqrt{n}\,\overline{X}$ und

$\sum_{i=1}^n (X_i-\overline{X})^2 = \sum_{i=1}^n X_i^2 - 2\underbrace{\sum_{i=1}^n X_i\overline{X}}_{n\overline{X}^2} + \underbrace{\sum_{i=1}^n \overline{X}^2}_{n\overline{X}^2} = X^\top X - Y_1^2 = (A^\top Y)^\top (A^\top Y) - Y_1^2 =$

$Y^\top AA^\top Y - Y_1^2 = Y^\top Y - Y_1^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2$

Beachte weiterhin, dass für $Z_1,\dots,Z_n$ i.i.d. mit $\mathcal{L}(Z_i) = N_{(0,1)}$ $\mathcal{L}(X_1,\dots,X_n) = \mathcal{L}(\sigma Z_1 + \mu,\dots,\sigma Z_n+\mu)$ gilt. □

4.1 Einstichprobentests bei Normalverteilungen

Im Folgenden: $X_1,\dots,X_n$ i.i.d. Zufallsvariablen mit $\mathcal{L}(X_i) = N_{(0,1)}$

$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$

$s_n^2(X) = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$

$t_{n,1-\alpha}$, $\mathcal{X}^2_{n,1-\alpha}$ bezeichne das α-Fraktil von t_n- bzw. $\mathcal{X}^2_n$-Verteilung.

Gauß-Test (Testen des Mittelwerts bei bekannter Varianz)

$H = \{\mu \le \mu_0\}$, $K = \{\mu > \mu_0\}$, $\sigma^2 = \sigma_0^2$ bekannt.

$$\varphi(x) = \begin{cases} 1 & > \\ \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma_0} & \Phi^{-1}(1-\alpha) \\ 0 & < \end{cases}$$

ist gleichmäßig bester Test zum Niveau α. (Beispiel 2.7)

t-Test (Testen des Mittelwerts bei unbekannter Varianz)

$\Theta = \mathbb{R}\times(0,\infty)$, $\vartheta = (\mu,\sigma^2)\in\Theta$

$H = \{(\mu,\sigma^2) : \mu\le\mu_0, \sigma^2>0\}$, $K = \{(\mu,\sigma^2) : \mu>\mu_0, \sigma^2>0\}$

$$\varphi(x) = \begin{cases} 1 & > \\ \sqrt{n}\frac{\overline{x}-\mu_0}{s_n(x)} & t_{n-1,1-\alpha} \\ 0 & < \end{cases}$$

ist $\{(\mu_0,\sigma^2) : \sigma^2>0\}$-ähnlich (vgl. Satz 4.2 4.)

$\mathcal{X}^2$-Test (Testen einer Varianz bei bekanntem oder unbekanntem Mittelwert)

Mittelwert μ_0 bekannt:

$\Theta = (0,\infty)$, $H = \{\sigma\le\sigma_0\}$, $K = \{\sigma>\sigma_0\}$, $\mu_0\in\mathbb{R}$ bekannt.

Die λ^n-Dichte von $N^n_{(\mu_0,\sigma^2)}$ ist

$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}\right)$, $x\in\mathbb{R}^n$

Also ist $Q = \{N^n_{(\mu_0,\sigma^2)} : \sigma>0\}$ Exponentialfamilie mit $q(\sigma) = -\frac{1}{2\sigma^2}$, $T(x) = \sum_{i=1}^n (x_i-\mu_0)^2$

Nach Satz 3.11 besitzt Q einen monotonen Dichtequotienten in T, da $q : (0,\infty)\mapsto\mathbb{R}$ streng monoton wachsend ist.

Nach Satz 3.6 ist der gleichmäßig beste Test φ zum Niveau α von der Form

$$\varphi(x) = \begin{cases} 1 & > \\ T(x) & c_\alpha \\ 0 & < \end{cases}$$

wobei c_α so gewählt sei, dass $E_{(\mu_0,\sigma_0^2)}\varphi = \alpha$

Seien $X_1, \ldots, X_n$ i.i.d. mit $\mathcal{L}(X_i) = N_{(\mu_0,\sigma_0^2)}$, dann $\mathcal{L}((X_i - \mu_0)/\sigma_0) = N_{(0,1)}$. Also:

$\sum_{i=1}^n \frac{(X_i-\mu_0)^2}{\sigma_0^2} = \sum_{i=1}^n (\frac{X_i-\mu_0}{\sigma_0})^2 \sim \mathcal{X}_n^2$ und damit

$$\varphi(x) = \begin{cases} 1 & > \\ \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sigma_0^2} & \mathcal{X}^2_{n,1-\alpha} \\ 0 & < \end{cases}$$

Mittelwert μ_0 unbekannt:

$\Theta = \mathbb{R} \times (0,\infty)$, $H = \{(\mu,\sigma^2) : \sigma \leq \sigma_0, \mu \in \mathbb{R}\}$, $K = \{(\mu,\sigma^2) : \sigma > \sigma_0, \mu \in \mathbb{R}\}$

Wähle Chiquadrat-Test

$$\varphi(x) = \begin{cases} 1 & > \\ \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\sigma_0^2} & \mathcal{X}^2_{n-1,-\alpha} \\ 0 & < \end{cases}$$

Nach Satz 4.2 gilt: $E_{(\mu,\sigma_0^2)}\varphi = \alpha$, da $\sum_{i=1}^n (X_i - \overline{X})^2/\sigma^2 \sim \mathcal{X}^2_{n-1}$, falls $\mathcal{L}(X_1, \ldots, X_n) = N^n_{(\mu,\sigma_0^2)}$

φ ist $\{(\mu,\sigma_0^2) : \mu \in \mathbb{R}\}$-ähnlich.

Zweiseitiger Gauß-Test

$H = \{\mu = \mu_0\}$ gegen $K = \{\mu \neq \mu_0\}$, $\sigma^2 = \sigma_0^2$ bekannt.

$$\varphi(x) = \begin{cases} 1 & > \\ \sqrt{n}\frac{|\overline{x}-\mu_0|}{\sigma_0} & \Phi^{-1}(1-\frac{\alpha}{2}) \\ 0 & < \end{cases}$$

Zweiseitiger t-Test

$H = \{(\mu_0,\sigma^2) : \sigma > 0\}$, $K = \{(\mu,\sigma^2) : \mu \neq \mu_0, \sigma > 0\}$

$$\varphi(x) = \begin{cases} 1 & > \\ \sqrt{n}\frac{|\overline{x}-\mu_0|}{s_n(x)} & t_{n-1,1-\frac{\alpha}{2}} \\ 0 & < \end{cases}$$

4.2 Zweistichprobentests

Ein Medikament soll auf Wirksamkeit untersucht werden. Es werden $n+m$ Personen zufällig ausgewählt. n Personen erhalten eine Behandlung mit dem Medikament, m Personen erhalten ein Placebo.

$X_1, \ldots, X_n$ Behandelte Personen

$Y_1, \ldots, Y_m$ Kontrollgruppe

Im Folgenden seien $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ unabhängig.

Gauß-Test (Vergleich zweier Mittelwerte bei bekannter identischer Varianz)

$\mathcal{Q} = \{N^n_{(\mu_x,\sigma_0^2)} \times N^m_{(\mu_y,\sigma_0^2)} : \mu_x, \mu_y \in \mathbb{R}\}$, σ_0^2 bekannt.

$\mathcal{L}(X_i) = N_{(\mu_x,\sigma_0^2)}$, $\mathcal{L}(Y_i) = N_{(\mu_y,\sigma_0^2)}$

$H = \{\mu_x \leq \mu_y\},\ K = \{\mu_x > \mu_y\}$
Betrachte den Test

$$\varphi(x,y) = \begin{cases} 1 & > \\ \sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{x}-\overline{y}}{\sigma_0} & \Phi^{-1}(1-\alpha) \\ 0 & < \end{cases}$$

Es gilt:
$\sqrt{n}\frac{\overline{X}-\mu_x}{\sigma_0} \sim N_{(0,1)},\ \sqrt{m}\frac{\overline{Y}-\mu_y}{\sigma_0} \sim N_{(0,1)}$
Also:
$\mathcal{L}(\sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{X}-\mu_x-(\overline{Y}-\mu_y)}{\sigma_0}) = N_{(0,\frac{m}{m+n})} * N_{(0,\frac{n}{m+n})} = N_{(0,\frac{m+n}{m+n})} = N_{(0,1)}$ (wobei „$*$“ die Faltung bezeichnet)
Die Gütefunktion von φ lautet
$E_{N^n_{(\mu_x,\sigma_0^2)}\times N^m_{(\mu_y,\sigma_0^2)}}(\varphi) = P\{\sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{X}-\overline{Y}-\mu_x+\mu_y}{\sigma_0} > \Phi^{-1}(1-\alpha) - \sqrt{\frac{n\cdot m}{n+m}}\frac{\mu_x-\mu_y}{\sigma_0}\} = 1-\Phi(\Phi^{-1}(1-\alpha) - \sqrt{\frac{n\cdot m}{n+m}}\frac{\mu_x-\mu_y}{\sigma_0}) = \begin{cases} < \alpha & \mu_x < \mu_y \\ \alpha & \mu_x = \mu_y \\ > \alpha & \mu_x > \mu_y \end{cases}$

t-Test (Vergleich zweier Mittelwerte bei unbekannter, identischer Varianz)
$\mathcal{Q} = \{N^n_{(\mu_x,\sigma^2)} \times N^m_{(\mu_y,\sigma^2)} : \mu_x, \mu_y \in \mathbb{R}, \sigma > 0\}$
Teste $H = \{\mu_x \leq \mu_y, \sigma > 0\}$ gegen $K = \{\mu_x > \mu_y, \sigma > 0\}$
Setze $S^2_{X,n} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2,\ S^2_{Y,m} = \frac{1}{m-1}\sum_{i=1}^m (Y_i - \overline{Y})^2,\ S^2_{n,m} = \frac{(n-1)S^2_{X,n}+(m-1)S^2_{Y,m}}{n+m-2}$
Der Test

$$\varphi(x,y) = \begin{cases} 1 & > \\ \sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{x}-\overline{y}}{S_{n,m}} & t_{n+m-2,1-\alpha} \\ 0 & < \end{cases}$$

heißt **Zweistichproben t-Test**.
Mit Lemma 4.4 folgt

$$E\varphi(X,Y) \begin{cases} < \alpha & \mu_x < \mu_y \\ = \alpha & \mu_x = \mu_y \\ > \alpha & \mu_x > \mu_y \end{cases}$$

Lemma 4.4

$$\mathcal{L}\left(\sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{X}-\overline{Y}-(\mu_x+\mu_y)}{S_{n,m}}\right) = t_{n+m-2}$$

Beweis: Nach Satz 4.2 1.: $\overline{X}, S^2_{X,n}, \overline{Y}, S^2_{Y,m}$ und damit auch $V_{n,m} = \sqrt{\frac{n\cdot m}{n+m}}\frac{\overline{X}-\overline{Y}-(\mu_x+\mu_y)}{\sigma}$ und $W_{n,m} = \frac{(n-1)S^2_{X,n}}{\sigma^2} + \frac{(m-1)S^2_{Y,m}}{\sigma^2}$ unabhängig.
$V_{n,m} \sim N_{(0,1)}$ (Gauß-Test),
$W_{n,m} \sim \mathcal{X}^2_{n-1} * \mathcal{X}^2_{m-1} = \mathcal{X}^2_{n+m-2}$
Mit der Definition der t-Verteilung folgt die Behauptung. □

F-Test (Vergleich zweier Varianzen bei bekannten oder unbekannten Mittelwerten)
Mittelwert bekannt:

$\mathcal{Q} = \{N^n_{(\mu_x,\sigma^2)} \times N^n_{(\mu_y,\tau^2)} : \sigma,\tau > 0\}$, μ_x, μ_y bekannt.
$H = \{\sigma \le \tau\}$, $K = \{\sigma > \tau\}$
Test:

$$\varphi(x,y) = \begin{cases} 1 & > \\ \frac{\frac{1}{n}\sum_{i=1}^n (x_i-\mu_x)^2}{\frac{1}{m}\sum_{i=1}^m (y_i-\mu_y)^2} & F_{n,m,1-\alpha} \\ 0 & < \end{cases}$$

Mittelwert unbekannt:
$\mathcal{Q} = \{N^n_{(\mu_x,\sigma^2)} \times N^n_{(\mu_y,\tau^2)} : \mu_x,\mu_y \in \mathbb{R}, \sigma,\tau > 0\}$
$H = \{\sigma \le \tau, \mu_x,\mu_y \in \mathbb{R}\}$, $K = \{\sigma > \tau, \mu_x,\mu_y \in \mathbb{R}\}$

$$\varphi(x) = \begin{cases} 1 & > \\ \frac{\frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2}{\frac{1}{m-1}\sum_{i=1}^m (y_i-\overline{y})^2} & F_{n-1,m-1,1-\alpha} \\ 0 & < \end{cases}$$

5 Optimale Konfidenzbereiche

Im Folgenden betrachten wir Aussagen der Form:
ϑ liegt mit Wahrscheinlichkeit $\ge \beta$ im Intervall $[t(x),\infty)$ oder $[t_1(x),t_2(x)]$,
wobei die Bereiche oder Intervalle auf der Beobachtung x (z.B. $\in \mathbb{R}^n$) basieren.

Beispiel 5.1 Sei $\mathcal{Q} = \{N_{(\mu,\sigma_0^2)} : \mu \in \mathbb{R}\}$, $\sigma_0 > 0$ fest.
Sei $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $x \in \mathbb{R}^n$. Dann:
$N^n_{(\mu,\sigma_0^2)}(\{x \in \mathbb{R}^n : \sqrt{n}\frac{|\overline{x}-\mu|}{\sigma_0} \le \Phi^{-1}(1-\frac{\alpha}{2})\}) = 1-\alpha = N_{(0,1)}([-\Phi^{-1}(1-\frac{\alpha}{2}), \Phi^{-1}(1-\frac{\alpha}{2}]])$
Offensichtlich
$\sqrt{n}\frac{|\overline{x}-\mu|}{\sigma_0} \le \Phi^{-1}(1-\frac{\alpha}{2})$
$\Leftrightarrow -\Phi^{-1}(1-\frac{\alpha}{2}) \le \sqrt{n}\frac{\mu-\overline{x}}{\sigma_0} \le \Phi^{-1}(1-\frac{\alpha}{2})$
$\Leftrightarrow \overline{x} - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}) \le \mu \le \overline{x} + \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2})$
$\Leftrightarrow \mu \in B(x) := [\overline{x} - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}), \overline{x} + \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2})]$
Beachte: B hängt nicht von μ ab.
Also: $N^n_{(\mu,\sigma_0^2)}(\{x \in \mathbb{R}^n : \mu \in B(x)\}) = 1-\alpha \quad \forall \mu \in \mathbb{R}$

Interpretation:
Das Intervall $B(x)$ zu einer Beobachtung x enthält mit Wahrscheinlichkeit $1-\alpha$ den wahren Parameter μ.

Im folgenden sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(S,\mathcal{B})$ und $\kappa : \Theta \mapsto K$ ein Funktional.

Beispiel 5.2 $(S,\mathcal{B}) = (\mathbb{R},\mathbb{B})$
$\vartheta \mapsto \int x\, dQ_\vartheta(x)$ Mittelwert von Q_ϑ
$\vartheta \mapsto F_\vartheta^{-1}(\frac{1}{2})$ Median, wobei F_ϑ Verteilungsfunktion zu Q_ϑ sei.

Definition 5.3 *Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(S,\mathcal{B})$ und $\kappa : \Theta \mapsto K$ ein Funktional.*

1. *Eine Abbildung* $B : S \mapsto \mathbb{P}(K)$ *heißt* ***Konfidenzbereich****, falls gilt*

$$\forall \kappa' \in K \quad \{x \in S : \kappa' \in B(x)\} \in \mathcal{B}$$

2. B *heißt Konfidenzbereich* ***zum Niveau*** $1-\alpha$*, falls gilt*

$$\forall \vartheta \in \Theta : \quad Q_\vartheta(\{x \in S : \kappa(\vartheta) \in B(x)\}) \geq 1-\alpha$$

Wir wollen einen Konfidenzbereich zum Niveau $1-\alpha$ als optimal ansehen, wenn „falsche" Parameterwerte mit möglichst geringer Wahrscheinlichkeit überdeckt werden. Welche Parameterwerte als falsch angesehen werden, hängt von der Problemstellung ab.
Zu $\kappa_0 \in K$ bezeichne $K_{\kappa_0} \subset K$, $\kappa_0 \notin K_{\kappa_0}$ die Menge der falschen Parameterwerte.

Beispiel 5.4 1. „Der Mittelwert der Verteilung von Q_ϑ ist mindestens gleich $\kappa(\vartheta)$" $\longrightarrow$ $K_{\kappa(\vartheta)} = (-\infty, \kappa(\vartheta))$

2. Zweiseitige Konfidenzintervalle wie in Beispiel 5.1 $\longrightarrow K_{\kappa(\vartheta)} = K \setminus \{\kappa(\vartheta)\}$

Definition 5.5 *Sei* $\{K_{\kappa'} : \kappa' \in K\}$ *eine Familie von falschen Parameterwerten.*
Ein Konfidenzbereich B^* *zum Niveau* α *heißt* ***gleichmäßig bester Konfidenzbereich zum Niveau*** α*, falls gilt:*

$$\forall \vartheta \in \Theta \ \forall \kappa' \in K_{\kappa(\vartheta)} \quad Q_\vartheta(\{x : \kappa' \in B^*(x)\}) = \inf_B Q_\vartheta(\{x : \kappa' \in B(x)\})$$

wobei das Infimum über alle Konfidenzbereiche B *zum Niveau* α *gebildet wird.*

Zur Erinnerung: C heißt gleichmäßig bester kritischer Bereich zum Niveau α, falls $\varphi = 1_C$ gleichmäßig bester Test zum Niveau α ist.

Sei nun B ein Konfidenzbereich zum Niveau $1-\alpha$, $\vartheta \in \Theta$, $K_{\kappa(\vartheta)} \subset K$ falsche Parameterwerte. Dann
$1-\alpha \leq Q_\vartheta(\{x : \kappa(\vartheta) \in B(x)\}) = 1 - Q_\vartheta(\{x : \kappa(\vartheta) \notin B(x)\}) = 1 - Q_\vartheta(C_{\kappa(\vartheta)})$,
wobei $\kappa' \in K : \quad C_{\kappa'} := \{x : \kappa' \notin B(x)\}$
Offenbar
$\forall \vartheta' \in \Theta, \kappa(\vartheta') = \kappa' \quad Q_{\vartheta'}(C_{\kappa'}) \leq \alpha$ $\qquad (*)$
und $Q_\vartheta(C_{\kappa'})$ möglichst groß, falls $\kappa' \in K_{\kappa(\vartheta)}$
$\Leftrightarrow \vartheta \in \widetilde{K}_{\kappa'} := \{\vartheta_0 : \kappa' \in K_{\kappa(\vartheta_0)}\}$
Betrachte $C_{\kappa'}$ als kritischen Bereich zum Niveau α für das Testproblem $\{\vartheta : \kappa(\vartheta) = \kappa'\}$ gegen $\widetilde{K}_{\kappa'}$.
Ist umgekehrt $C_{\kappa'}$ kritischer Bereich zum Niveau α für $\{\vartheta : \kappa(\vartheta) = \kappa'\}$ gegen $\{\vartheta : \kappa' \in K_{\kappa(\vartheta)}\}$, $\kappa' \in K$, so ist $B(x) = \{\kappa' \in K : x \notin C_{\kappa'}\}$ ein Konfidenzbereich zum Niveau $1-\alpha$, denn
$Q_\vartheta\{x : \kappa(\vartheta) \in B(x)\} = 1 - Q_\vartheta\{x : \kappa(\vartheta) \notin B(x)\} = 1 - Q_\vartheta(C_{\kappa(\vartheta)}) \geq 1-\alpha$ $\qquad (**)$

Satz 5.6 *1. Sei* B *ein Konfidenzbereich zum Niveau* $1-\alpha$*. Dann gilt für alle* $\vartheta_1 \in \Theta, \kappa' \in K_{\kappa(\vartheta_1)}$*:*

$$C_{\kappa'} := \{x \in S : \kappa' \notin B(x)\}$$

ist kritischer Bereich zum Niveau α *für* $\{\vartheta \in \Theta : \kappa(\vartheta) = \kappa'\}$ *gegen* $\{\vartheta_1\}$*.*

2. *Für alle* $\vartheta_0 \in \Theta$ *sei* $C_{\kappa(\vartheta_0)}$ *ein kritischer Bereich mit* $Q_{\vartheta_0}(C_{\kappa(\vartheta_0)}) \leq \alpha$. *Dann wird durch*

$$B(x) := \{\kappa' \in K : x \notin C_{\kappa'}\}$$

ein Konfidenzbereich zum Niveau $1-\alpha$ *definiert.*

3. B^* *ist ein gleichmäßig bester Konfidenzbereich zum Niveau* $1-\alpha$, *falls*

$$C^*_{\kappa'} := \{x : \kappa' \notin B^*(x)\}$$

gleichmäßig bester kritischer Bereich zum Niveau $1-\alpha$ *für* $\{\vartheta \in \Theta : \kappa(\vartheta) = \kappa'\}$ *gegen* $\{\vartheta : \kappa' \in K_{\kappa(\vartheta)}\}$ *ist.*

Beweis:
1. $C_{\kappa'} \in \mathcal{B}$ und $Q_\vartheta(C_{\kappa'}) \leq \alpha \ \forall \kappa(\vartheta) = \kappa'$ folgt aus (*).
2. folgt aus (**)
3. Sei B wie in 1. Dann $\forall \vartheta_1 \in \Theta \ \forall \kappa' \in K_{\kappa(\vartheta_1)}$:
$Q_{\vartheta_1}(\{x : \kappa' \in B^*(x)\}) = 1 - \underbrace{Q_{\vartheta_1}(C^*_{\kappa'})}_{\geq Q_{\vartheta_1}(C_{\kappa'})} \leq 1 - Q_{\vartheta_1}(C_{\kappa'}) = Q_{\vartheta_1}(\{x : \kappa' \in B(x)\})$ □

Satz 5.7 $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ *habe einen monotonen Dichtequotienten in* T, *so dass für alle* $\vartheta \in \Theta$ TQ_ϑ *eine stetige Verteilung ist.*[2]
Sei $\kappa : \Theta \mapsto K = \Theta$, $\kappa(\vartheta) = \vartheta$ *und* $K_\vartheta = \{\vartheta' \in \Theta : \vartheta' < \vartheta\} \ \forall \vartheta \in \Theta$.
Dann existiert ein gleichmäßig bester Konfidenzbereich B^* *zum Niveau* α *mit*

$$B^*(x) = \{\vartheta \in \Theta : T(x) \leq c_\vartheta\}$$

wobei c_ϑ *bestimmt wird durch* $Q_\vartheta(\{x : T(x) > c_\vartheta\}) = \alpha$.

Beweis: Sei $\vartheta_1 \in \Theta, \vartheta_0 \in K_{\vartheta_1}$, d.h. $\vartheta_0 < \vartheta_1$.
Gemäß Satz 3.6 existiert ein gleichmäßig bester Test φ_{ϑ_0} zum Niveau α mit

$$\varphi_{\vartheta_0}(x) = \begin{cases} 1 & > \\ T(x) & c_{\vartheta_0} \\ 0 & \leq \end{cases}$$

für $\{\vartheta_0\}$ gegen $\{\vartheta' \in \Theta : \vartheta' > \vartheta_0\}$. Dann ist $C_{\vartheta_0} = \{x \in S : T(x) > c_{\vartheta_0}\}$ gleichmäßig bester kritischer Bereich zum Niveau α für dieses Testproblem und nach Satz 5.6
$\{\vartheta \in \Theta : x \notin C_\vartheta\} = \{\vartheta \in \Theta : T(x) \leq c_\vartheta\} = B^*$ gleichmäßig bester Konfidenzbereich zum Niveau $1-\alpha$. □

Beispiel 5.8 $\mathcal{Q} = \{N^n_{(\mu,\sigma_0^2)} : \mu \in \mathbb{R}\}, \sigma_0 > 0$ fest, $\kappa(\mu) = \mu$ Funktional.

1. *Untere Konfidenzschranke zum Niveau* $1-\alpha$
$K_\mu := (-\infty, \mu)$ falsche Parameterwerte
$\mathcal{Q}$ hat monotonen Dichtequotienten in $T(x) = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$

[2] $TQ_\vartheta = Q^T_\vartheta = Q_\vartheta\{T \leq \cdot\}$

$TN^n_{(\mu,\sigma_0^2)}$ ist stetige Verteilung
Also nach Satz 5.7 $B^*(x) = \{\mu \in \mathbb{R} : \overline{x} \leq c_\mu\}$ gleichmäßig bester Konfidenzbereich, wobei $\alpha = N^n_{(\mu,\sigma_0^2)}(\{x \in \mathbb{R}^n : \overline{x} > c_\mu\}) = N^n_{(\mu,\sigma_0^2)}(\{x \in \mathbb{R}^n : \sqrt{n}\frac{\overline{x}-\mu}{\sigma_0} > \sqrt{n}\frac{c_\mu-\mu}{\sigma_0}\}) = 1 - \Phi(\sqrt{n}\frac{c_\mu-\mu}{\sigma_0}) \Leftrightarrow \sqrt{n}\frac{c_\mu-\mu}{\sigma_0} = \Phi^{-1}(1-\alpha) \Leftrightarrow c_\mu = \mu + \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha)$
Also $B^*(x) = \{\mu \in \mathbb{R} : \mu \geq \overline{x} - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha)\} = [\overline{x} - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha), \infty)$

2. *Konfidenzintervall*
$K_\mu = \mathbb{R} \setminus \{\mu\}$
Zweiseitiger Gauß-Test zum Niveau α für $\{\mu_1\}$ gegen $\{\mu : \mu_1 \in K_\mu\} = \mathbb{R} \setminus \{\mu_1\}$
$C_{\mu_1} = \{x \in \mathbb{R}^n : \sqrt{n}\frac{|\overline{x}-\mu_1|}{\sigma_0} > \Phi^{-1}(1-\frac{\alpha}{2})\}$
Dann
$B(x) = \{\mu \in \mathbb{R} : x \notin C_\mu\} = \{\mu \in \mathbb{R} : \sqrt{n}\frac{|\overline{x}-\mu|}{\sigma_0} \leq \Phi^{-1}(1-\frac{\alpha}{2})\} = [\overline{x} - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}), \overline{x} + \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2})]$
(vergleiche Beispiel 5.1)

Kapitel II

Schätztheorie

6 Vorbemerkungen

Zur Erinnerung: **Bedingte Erwartung, bedingte Verteilung**
$(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum, $X : (\Omega, \mathcal{A}, P) \mapsto (\mathbb{R}^n, \mathbb{B}^n)$, $Y : (\Omega, \mathcal{A}, P) \mapsto (T, \mathcal{C})$, $\mathcal{A}_0 = Y^{-1}(\mathcal{C}) = \sigma(Y)$

$E(X|Y)$ **bedingte Erwartung von** X **gegeben** Y , falls $E(X|Y) : (\Omega, \mathcal{A}_0) \mapsto (\mathbb{R}, \mathbb{B})$ und $\forall A \in \mathcal{A}_0 : \; E(1_A E(X|Y)) = E(1_A X)$

$E(X|Y = y)$ **bedingte Erwartung von** X **gegeben** $Y = y$, falls
$E(X|Y = \cdot) : (T, \mathcal{C}) \mapsto (\mathbb{R}, \mathbb{B})$ und $\forall C \in \mathcal{C} : \; E(1_C(Y)X) = \int_C E(X|Y = y) \; d\mathcal{L}(Y)(y)$

$P(X \in B|Y = y)$ **bedingte Wahrscheinlichkeit von** X **gegeben** $Y = y$, falls $\forall B \in \mathbb{B}^n$:
$P(X \in B|Y = y) = E(1_B(X)|Y = y)$,
d.h. $P\{X \in B, Y \in C\} = \int_C P(X \in B|Y = y) \; d\mathcal{L}(Y)(y)$

$P(X \in \cdot\,|Y = y)$ **bedingte Verteilung oder reguläre bedingte Wahrscheinlichkeit** , falls $\forall y \in T \; P(X \in \cdot\,|Y = y)$ Wahrscheinlichkeitsmaß und $P(X \in \cdot\,|Y = y)$ bedingte Wahrscheinlichkeit ist.

Im Folgenden $P(X \in \cdot\,|Y = y)$ bedingte Verteilung von X gegeben $Y = y$.

Lemma 6.1

$$\mathcal{L}(Y, X) = K\mathcal{L}(Y)$$

wobei $K(\cdot\,|y) = \varepsilon_y \times P(X \in \cdot\,|Y = y), \; K\mathcal{L}(Y)(B) = \int K(B|y) \; d\mathcal{L}(Y)(y)$

Beweis: $K(\cdot\,|y)$ ist Wahrscheinlichkeitsmaß und $\forall B_1 \in \mathcal{C}, B_2 \in \mathbb{B}^n : \; K(B_1 \times B_2|y) = 1_{B_1}(y) P(X \in B_2|Y = y)$
Also
$K\mathcal{L}(Y)(B_1 \times B_2) = \int K(B_1 \times B_2|y) \; d\mathcal{L}(Y)(y) = \int 1_{B_1}(y) P(X \in B_2|Y = y) \; d\mathcal{L}(Y)(y) = P\{X \in B_2, Y \in B_1\} = \mathcal{L}(Y, X)(B_1 \times B_2)$ □

Lemma 6.2 $(n = 1)$

$$E(X|Y = y) = \int xP(X \in dx|Y = y) \quad \mathcal{L}(Y) - f.s.$$

Beweis: Mit $g(y, x) = x$: $\int xP(X \in dx|Y = y) = \int g\ dK(\cdot|y)$.
Also
$\int 1_{B_1}(y) \int xP(X \in dx|Y = y)d\mathcal{L}(Y)(y) = \int(\int 1_{B_1\times\mathbb{R}} g\ dK(\cdot|y))d\mathcal{L}(Y)(y) = \int 1_{B_1\times\mathbb{R}} g\ d\underbrace{K\mathcal{L}(Y)}_{\mathcal{L}(Y,X)} =$
$\int 1_{B_1}(y)x\ d\mathcal{L}(Y, X)(y, x) = E(1_{B_1}(Y)X)$ □

Lemma 6.3 *$gP(X \in \cdot|Y = y)$ ist bedingte Verteilung von $g(X)$ gegeben $Y = y$.*

Beweis: $(gP(X \in \cdot|Y = y))(B) = P(X \in \cdot|Y = y)(g^{-1}(B)) = P(X \in g^{-1}(B)|Y = y) = E(1_{g^{-1}(B)}(X)|Y = y) = E(1_B(g(X))|Y = y) \quad \mathcal{L}(Y)$-f.s.
Also $P(g(X) \in \cdot|Y = y) = gP(X \in \cdot|Y = y)$ □

Lemma 6.4 *$\varepsilon_y \times P(X \in \cdot|Y = y)$ ist bedingte Verteilung von (Y, X) gegeben $Y = y$.*

Beweis: $P\{(Y, X) \in B_1 \times D, Y \in B_2\} = P\{Y \in B_1 \cap B_2, X \in D\} = \int 1_{B_1\cap B_2}(y)P(X \in D|Y = y)d\mathcal{L}(Y)(y) = \int_{B_2}(\varepsilon_y \times P(X \in \cdot|Y = y))(B_1 \times D)d\mathcal{L}(Y)(y)$
Also $P((Y, X) \in \cdot|Y = y) = \varepsilon_y \times P(X \in \cdot|Y = y)$ □

Lemma 6.5

$$E(g(Y, X)|Y = y) = \int g(y, x)P(X \in dx|Y = y)$$

Beweis: $E(g(X, Y)|Y = y) = \int zP(g(Y, X) \in dz|Y = y) = \int z(gP((Y, X) \in \cdot|Y = y)(z) = \int g(z_1, z_2)P((Y, X) \in d(z_1, z_2)|Y = y) = \int g(z_1, z_2)d(\varepsilon_y \times P(X \in \cdot|Y = y))(z_1, z_2) =$
$\int(\underbrace{\int g(z_1, z_2)d\varepsilon_y(z_1)}_{g(y,z_2)})P(X \in dz_2|Y = y)$ □

Im Folgenden:
$(S, \mathcal{B})$ messbarer Raum
Θ Parametermenge
$\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B})$.

Bezeichnung 6.6 *$\mathcal{Q}$ heißt **statistisches Experiment**.*
*Ist $\Theta \subset \mathbb{R}^n$, so spricht man von einem **parametrischen statistischen Experiment**, ansonsten von einem nichtparametrischen Experiment.*
Allgemein kann Θ auch eine Teilmenge der Wahrscheinlichkeitsmaße auf $(S, \mathcal{B})$ sein.

Definition 6.7 *Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B})$, D eine Menge.*

1. *Eine Abbildung $g : \mathcal{Q} \mapsto D$ oder $g : \Theta \mapsto D$ heißt* ***statistisches Funktional****.*
2. *Ist $(D, \mathcal{D})$ ein messbarer Raum, so heißt eine Abbildung $T : (S, \mathcal{B}) \mapsto (D, \mathcal{D})$* ***Schätzer****.*
3. *T ist erwartungstreu für g, falls*

$$\forall \vartheta \in \Theta \quad E_\vartheta T = g(\vartheta) \quad (bzw.\ g(Q_\vartheta))$$

Definition 6.8 *Eine Abbildung $T : (S, \mathcal{B}) \mapsto (S', \mathcal{B}')$ heißt* ***Statistik****.*

Definition 6.9 *K heißt* ***Markov-Kern****, wenn*

1. $\forall A \in \mathcal{B}: \quad K(A|\cdot) : (S', \mathcal{B}') \mapsto ([0,1], \mathbb{B}|_{[0,1]})$
2. $\forall t \in S': \quad K(\cdot\,|t)$ *Wahrscheinlichkeitsmaß auf* $(S, \mathcal{B})$

7 Suffizienz

Bei Verteilungsklassen mit monotonem Dichtequotienten T und einseitigen Testproblemen haben wir gesehen, dass der beste Test nach T faktorisiert. Bei der Suche nach optimalen Tests kann man sich auf solche beschränken, die die Form

$$\varphi = h \circ T$$

haben. Im folgenden werden wir sehen, dass analoge Resultate auch für gewisse Schätzprobleme gelten.

Beispiel 7.1 Beobachtet man n unabhängige Wiederholungen eines Versuchs mit Werten in $\{0,1\}$ ($1 \Leftrightarrow$ Erfolg, $0 \Leftrightarrow$ kein Erfolg), so wird man als Modell
$\mathcal{Q} = \{B^n_{(1,p)} : p \in [0,1]\}, \quad (S, \mathcal{B}) = (\{0,1\}, \mathbb{P}(\{0,1\}^n), \quad \Theta = [0,1], \quad x = (x_1, \ldots, x_n) \in S$
wählen.
Man wird vermuten, dass die Statistik $T(x) = \sum_{i=1}^{n} x_i$ die gleiche Information wie x enthält.
Betrachte hierzu die bedingte Verteilung von $B^n_{(1,p)}$ gegeben $T = k$ bzw. $P(X \in \cdot\,|T(x) = k)$, wobei $\mathcal{L}(X) = B^n_{(1,p)}$:
$P(X = x|T(x) = k) = \frac{P(X=x, T(x)=k)}{P(T(x)=k)} = \frac{P(X=x)}{P(T(x)=k)} = \frac{p^{\Sigma x_i} \cdot (1-p)^{n-\Sigma x_i}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{\binom{n}{k}}$ falls $T(x) = k$ und 0 sonst.
Offenbar ist die bedingte Verteilung von X gegeben $T(x)$ unabhängig von p.

Allgemein gilt für die bedingte Verteilung $Q_\vartheta(\cdot\,|T = t)$

$$Q_\vartheta(A) = E_\vartheta 1_A = \int E_\vartheta(1_A|T = t) dTQ_\vartheta(t) = \int Q_\vartheta(A|T = t) dTQ_\vartheta(t)$$

Ist jetzt $Q_\vartheta(A|T=t)$ unabhängig von ϑ wie in Beispiel 7.1, so schreiben wir $Q_\cdot(A|T=t)$ und es gilt

$$Q_\vartheta(A) = \int Q_\cdot(A|T=t)dTQ_\vartheta(t)$$

Offenbar steckt die gesamte Information über ϑ bzw. Q_ϑ in TQ_ϑ, da $Q_\cdot(A|T=t)$ nicht von ϑ abhängt.
Ist $Q_\cdot(A|T=t)$ eine bedingte Verteilung, also $K(A|t) = Q_\cdot(A|T=t)$ ein Markov-Kern, so lässt sich $\{Q_\vartheta : \vartheta \in \Theta\}$ aus $\{TQ_\vartheta : \vartheta \in \Theta\}$ (in Beispiel 7.1: $\{B^n_{(1,p)} : p \in [0,1]\}$ aus $\{B_{(n,p)} : p \in [0,1]\}$) durch ein zweistufiges Experiment reproduzieren: Ziehe Beobachtung t nach TQ_ϑ und dann Beobachtung x nach $K(\cdot|t)$.

Definition 7.2 *1. Sei $\mathcal{B}_0 \subset \mathcal{B}$ eine σ-Algebra.*
*$\mathcal{B}_0$ heißt **suffizient** (für $\mathcal{Q}$), falls für alle $A \in \mathcal{B}$ eine von ϑ unabhängige Version $E_\cdot(1_A|\mathcal{B}_0)$ der bedingten Erwartung $E_\vartheta(1_A|\mathcal{B}_0) = Q_\vartheta(A|\mathcal{B}_0)$ existiert.*

*2. Eine Statistik $T : (S,\mathcal{B}) \mapsto (S',\mathcal{B}')$ heißt **suffizient**, falls $\mathcal{B}_0 := T^{-1}(\mathcal{B}')$ suffizient ist.*

Interpretation: Ist T suffizient, so enthält das Experiment $\{TQ_\vartheta : \vartheta \in \Theta\}$ ebensoviel Information wie das Ausgangsexperiment.

Bemerkung 7.3 Ist $E_\vartheta|Z| < \infty$, $\vartheta \in \Theta$, für eine Statistik Z, so existiert eine von ϑ unabhängige Version $E_\cdot(Z|\mathcal{B}_0)$, falls $\mathcal{B}_0$ suffizient ist.

Satz 7.4 *Sei $T : (S,\mathcal{B}) \mapsto (S',\mathcal{B}')$ suffizient für $\mathcal{Q}$ und $Z : (S,\mathcal{B}) \mapsto (\mathbb{R}^n,\mathbb{B}^n)$.*
Dann existiert eine von ϑ unabhängige Version der bedingten Verteilung $Q_\cdot(Z \in \cdot|T=t)$.

Beweis: Sei $n=1$. Der Beweis für $n>1$ kann analog geführt werden.
Sei $\mathbb{Q}$ die Menge der rationalen Zahlen.
$\forall q \in \mathbb{Q}$ wähle $F(q|t) = E_\cdot(1_{(-\infty,q]}(Z)|T=t)$
Aufgrund der Eigenschaften bedingter Erwartungswerte gilt:
$B' = \bigcap_{\substack{q_1,q_2\in\mathbb{Q}\\ q_1<q_2}} \{t : F(q_1|t) \le F(q_2|t)\} \cap \{t : \lim_{q\to-\infty} F(q|t) = 0, \lim_{q\to\infty} F(q|t) = 1\} \cap \{t : \lim_{n\to\infty} F(q+\frac{1}{n}|t) = F(q|t),\ q \in \mathbb{Q}\} \in \mathcal{B}'$ und es gilt $Q_\vartheta\{T \in B'\} = 1\ \forall\vartheta \in \Theta$.
Für $t \in B'$ ist $\widetilde{F}(r|t) := \lim_{q\downarrow r} F(q|t)$ eine Verteilungsfunktion.
Sei $K(\cdot|t)$ die Verteilung mit Verteilungsfunktion $\widetilde{F}(\cdot|t)$, falls $t \in B'$ und $K(\cdot|t) = K(\cdot|t_0)$ für alle $t \notin B'$ und ein beliebiges $t_0 \in B'$.
Weiterhin sei $\mathcal{E} = \{(-\infty,q] : q \in \mathbb{Q}\}$; $\mathcal{E}$ ist abzählbar und ein $\cap$-stabiler Erzeuger von $\mathbb{B}$.
Dann gilt für alle $\vartheta \in \Theta$
$K(E|t) = F(q|t) = E_\cdot(1_{(-\infty,q]}(Z)|T=t) = Q_\vartheta(Z \in E|T=t) \quad TQ_\vartheta$-f.s. $\forall E = (-\infty,q] \in \mathcal{E}$[1] und damit (wegen $\mathcal{E}$ abzählbar)
$K(\cdot|t)|_\mathcal{E} = Q_\vartheta(Z \in \cdot|T=t)|_\mathcal{E} \quad TQ_\vartheta$-f.s.
$\Rightarrow K(\cdot|t) = Q_\vartheta(Z \in \cdot|T=t) \quad TQ_\vartheta$-f.s.
Aus der $(\mathcal{B}',\mathbb{B})$-Messbarkeit von $K(E|\cdot)$ $\forall E \in \mathcal{E}$ folgt mit Standardargumenten die Messbarkeit

[1] bedingte Verteilung von Z gegeben $T=t$ existiert nach Wahrscheinlichkeitstheorie

von $K(A|\cdot)$ $\forall A \in \mathbb{B}$. □

Korollar 7.5 *Sei* $f : (\mathbb{R}^n \times S', \mathbb{B}^n \times \mathcal{B}') \mapsto (\mathbb{R}, \mathbb{B})$ *mit* $\int |f(Z,T)|\, dQ_\vartheta < \infty$, $\vartheta \in \Theta$. *Dann*

$$\int f(Z,T)\, dQ_\vartheta = \iint f(z,t) Q_\cdot(Z \in dz | T = t)\, dTQ_\vartheta(t)$$

und

$$E_\cdot(f(Z,T)|T=t) = \int f(z,t) Q_\cdot(Z \in dz | T = t)$$

Beweis: Siehe Übung. □

Satz 7.6 (Neyman-Kriterium) *Sei* $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ *durch ein* σ*-endliches Maß* μ *dominiert* ($Q_\vartheta \ll \mu$ $\forall \vartheta \in \Theta$).
Eine Statistik T *ist genau dann suffizient, wenn eine nichtnegative Funktion* $h : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ *und nichtnegative Funktionen* $f_\vartheta : (S', \mathcal{B}') \mapsto (\mathbb{R}, \mathbb{B})$ *existieren, so dass gilt*

$$\forall \vartheta \in \Theta \quad f_\vartheta \circ T \cdot h \in \frac{dQ_\vartheta}{d\mu}$$

Beispiel 7.7 Siehe Beispiel 7.1:
$\frac{dB^n_{(1,p)}}{d\mu} = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i} = p^t(1-p)^{n-t} = f_p \circ T$ mit $T = \sum x_i$, μ Zählmaß.

Zum Beweis von Satz 7.6 benötigen wir zwei Hilfssätze und folgende

Definition 7.8 *Zwei Familien* $\mathcal{M}$ *und* $\mathcal{N}$ *von Maßen auf* $(S, \mathcal{B})$ *heißen* ***äquivalent*** *(i.Z.* $\mathcal{M} \approx \mathcal{N}$*), wenn*

$$(\forall \mu \in \mathcal{M} :\ \mu(A) = 0) \Longleftrightarrow (\forall \nu \in \mathcal{N} :\ \nu(A) = 0)$$

Satz 7.9 *Eine Familie von Wahrscheinlichkeitsmaßen ist durch ein* σ*-endliches Maß dominiert genau dann, wenn sie eine abzählbare äquivalente Teilfamilie besitzt.*

Beweis: Siehe Witting (1985), Satz 1.136. □

Bemerkung 7.10 Ist $\mathcal{Q}_0 = \{Q_{\vartheta_i} : \vartheta_i \in \Theta, i \in \mathbb{N}\} \subset \mathcal{Q}$ mit $\mathcal{Q}_0 \approx \mathcal{Q}$, so ist
$\nu := \sum_{i=1}^\infty c_i Q_{\vartheta_i} \approx \mathcal{Q}$, falls $c_i > 0$, $\sum c_i = 1$.

Satz 7.11 *Sei* $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ *dominiert durch ein* σ*-endliches Maß. Sei* $\nu = \sum_{i=1}^\infty c_i Q_{\vartheta_i}$ *äquivalent zu* $\mathcal{Q}$.
Dann ist T *suffizient für* $\mathcal{Q}$ *genau dann, wenn für alle* $\vartheta \in \Theta$ *eine nichtnegative Funktion* $f_\vartheta : (S', \mathcal{B}') \mapsto (\mathbb{R}, \mathbb{B})$ *existiert, so dass gilt:*

$$\forall \vartheta \in \Theta \quad f_\vartheta \circ T \in \frac{dQ_\vartheta}{d\nu}$$

Beweis:
$\Rightarrow$: $\mathcal{B}_0 = T^{-1}(\mathcal{B}')$ und $A \in \mathcal{B}$
Sei $E_\cdot(1_A|T)$ eine von ϑ unabhängige Version der bedingten Erwartung.
Wir zeigen: $E_\cdot(1_A|T) = E_\nu(1_A|T)$ ν-f.s. $\quad (*)$
Sei $B \in \mathcal{B}_0$. Dann
$\nu(A \cap B) = \sum_{i=1}^\infty c_i Q_{\vartheta_i}(A \cap B) = \sum_{i=1}^\infty c_i \int_B 1_A \, dQ_{\vartheta_i} = \sum_{i=1}^\infty c_i \int_B E_\cdot(1_A|T) dQ_{\vartheta_i} =$
$= \int_B E_\cdot(1_A|T) d(\sum_{i=1}^\infty c_i Q_{\vartheta_i}) = \int_B E_\cdot(1_A|T) d\nu \quad \Rightarrow (*)$
Wähle $\tilde{f}_\vartheta \in \frac{dQ_\vartheta|_{\mathcal{B}_0}}{d\nu|_{\mathcal{B}_0}}$. Nach dem Faktorisierungslemma (Wahrscheinlichkeitstheorie 19.2[2]) existiert $f_\vartheta : (S', \mathcal{B}') \mapsto (\mathbb{R}, \mathbb{B})$ mit $\tilde{f}_\vartheta = f_\vartheta \circ T$, also $f_\vartheta \circ T \in \frac{dQ_\vartheta|_{\mathcal{B}_0}}{d\nu|_{\mathcal{B}_0}}$.
Dann gilt: $Q_\vartheta(A) = \int E_\cdot(1_A|T) dQ_\vartheta|_{\mathcal{B}_0} \overset{(*)}{=} \int E_\nu(1_A|T) f_\vartheta \circ T \, d\nu|_{\mathcal{B}_0} = \int E_\nu(1_A f_\vartheta \circ T|T) d\nu|_{\mathcal{B}_0} =$
$\int E_\nu(1_A f_\vartheta \circ T|T) d\nu = \int 1_A f_\vartheta \circ T \, d\nu$
$\Rightarrow f_\vartheta \circ T \in \frac{dQ_\vartheta}{d\nu}$
$\Leftarrow$: Wir zeigen: $E_\nu(1_A|T)$ ist Version von $E_\vartheta(1_A|T) \quad \forall \vartheta$
Sei $B \in \mathcal{B}_0$
$Q_\vartheta(A \cap B) = \int 1_A 1_B f_\vartheta \circ T \, d\nu = \int E_\nu(1_A 1_B f_\vartheta \circ T|T) d\nu = \int 1_B E_\nu(1_A|T) f_\vartheta \circ T \, d\nu =$
$\int_B E_\nu(1_A|T) dQ_\vartheta$
$\Rightarrow E_\nu(1_A|T) = E_\vartheta(1_A|T)$ □

Damit können wir das Neyman-Kriterium beweisen:
Beweis: (zu Satz 7.6)
$\Rightarrow$: Sei T suffizient. Wähle abzählbare äquivalente Teilfamilie $\mathcal{Q}_0 \subset \mathcal{Q}$ gemäß Satz 7.9. Seien f_ϑ und ν wie in Satz 7.11 und $h = \frac{d\nu}{d\mu}$. (Beachte: $Q_{\vartheta_i} \ll \mu \Rightarrow \nu \ll \mu$). Dann gilt:
$Q_\vartheta(A) = \int_A f_\vartheta \circ T \, d\nu = \int_A f_\vartheta \circ T \cdot h \, d\mu$
$\Leftarrow$: Sei $A \in \mathcal{B}$. Dann gilt
$\nu(A) = \sum_{i=1}^\infty c_i Q_{\vartheta_i}(A) = \sum_{i=1}^\infty c_i \int_A f_{\vartheta_i} \circ T \cdot h \, d\mu = \int (\sum_{i=1}^\infty c_i f_{\vartheta_i} \circ T \cdot h) d\mu$
Weiter gilt $Q_\vartheta(A) = \int_A f_\vartheta \circ T \cdot h \, d\mu$ und mit $Q_\vartheta \ll \nu$
$\int_A \frac{f_\vartheta \circ T \cdot h}{\sum_{i=1}^\infty c_i f_{\vartheta_i} \circ T \cdot h} d\nu = \int_A f_\vartheta \circ T \cdot h \, d\mu = Q_\vartheta(A)$
Mit Satz 7.11 folgt die Behauptung. □

Beispiel 7.12 $\mathcal{Q} = \{B^n_{(1,p)} : p \in [0,1]\}$
Dann ist $T(x) = \sum_{i=1}^n x_i$ suffizient für p (vgl. Beispiel 7.1).
Es gilt
$B^n_{(1,p)}(\{x\}) = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i} = p^{T(x)}(1-p)^{n-T(x)} = f_p(T(x))$
$(f_p \circ T \in \frac{dB^n_{(1,p)}}{d(\varepsilon_0+\varepsilon_1)^n})$

Beispiel 7.13 $\mathcal{Q} = \{N^n_{(\mu,\sigma^2)} : \mu \in \mathbb{R}, \sigma^2 > 0\}$
$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i, \quad s_n^2(x) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$

[2] bzw. 19.9 in der Vorlesungsmitschrift ($\mu_\rho \varepsilon$)

Dann ist $T(x) = (\overline{x}, s_n^2(x))$ suffizient für (μ, σ^2):
$N^n_{(\mu,\sigma^2)}$ hat Dichte
$f_{(\mu,\sigma^2)}(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right)$
Es ist
$\sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (x_i-\overline{x})^2 + n(\overline{x}-\mu)^2 = (n-1)s_n^2(x) + n(\overline{x}-\mu)^2$
Also
$f_{(\mu,\sigma^2)}(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{(n-1)s_n^2(x)+n(\overline{x}-\mu)^2}{2\sigma^2}\right) = \tilde{f}_{(\mu,\sigma^2)} \circ T$

Beispiel 7.14 (Ordnungsstatistiken)
Zu $x \in \mathbb{R}^n$ bezeichne $x_{1:n}, \ldots, x_{n:n}$ die Ordnungsstatistiken, d.h. die der Größe nach geordneten Werte von $x_1, \ldots, x_n$. Dann ist $T(x) = (x_{1:n}, \ldots, x_{n:n})$ suffizient für $\mathcal{Q} = \{Q^n : Q \text{ Wahrscheinlichkeitsmaß auf } (\mathbb{R}, \mathbb{B})\}$

Beweis: Siehe Übung. □

8 Erwartungstreue Schätzer und Vollständigkeit

Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ und ein statistisches Funktional $g : \Theta \mapsto D$ gegeben.
Ist $D \subset \mathbb{R}$, so bietet sich an, die Qualität eines Schätzers h für g anhand der Funktion $\vartheta \mapsto E_\vartheta((h - g(\vartheta))^2)$ zu beurteilen.

Definition 8.1 *1. Sei $(D, \mathcal{D})$ ein Maßraum und $L(\vartheta, \cdot) : D \mapsto \overline{\mathbb{R}}$ für jedes $\vartheta \in \Theta$ eine messbare Abbildung mit*

$$\forall d \in D: \quad L(\vartheta, d) \geq 0 \text{ und } L(\vartheta, g(\vartheta)) = 0$$

Dann heißt L ***Verlustfunktion.***

2. Sei h ein (Punkt-)Schätzer für $g(\vartheta)$. Dann heißt

$$R(\vartheta, h) = E_\vartheta(L(\vartheta, h))$$

Risikofunktion.

Interpretation: $L(\vartheta, h(x))$ gibt den Verlust an, wenn bei Vorliegen von ϑ die Schätzung $h(x)$ aufgrund der Realisation x vorgenommen wird.

Beispiel 8.2 $D = \mathbb{R}$, $p > 0$
$L(\vartheta, d) := |g(\vartheta) - d|^p$
$p = 2$: $L(\vartheta, d) = (g(\vartheta) - d)^2$ **Gaußsche Verlustfunktion**

Im Folgenden: $D \subset \mathbb{R}$

Bemerkung 8.3 1. Für einen Schätzer h heißt

$$b(\vartheta, h) = E_\vartheta(h) - g(\vartheta)$$

bias (Verzerrung).

2. Bis auf triviale Fälle existiert i.A. kein Schätzer h, der die Risikofunktion gleichmäßig in ϑ minimiert. Wähle z.B. $\tilde{h}_{\vartheta_0} = g(\vartheta_0)$. Dann $R(\vartheta_0, \tilde{h}_{\vartheta_0}) = 0$.
Deshalb liegt es nahe, das Risiko in einer gewissen Teilklasse von Schätzern zu minimieren, z.B. in der Klasse der erwartungstreuen Schätzer, also $b(\vartheta, h) = 0$.

8.1 Konvexe Verlustfunktionen

Definition 8.4 *Eine Verlustfunktion heißt (strikt)* ***konvex****, falls $L(\vartheta, \cdot)$ für jedes $\vartheta \in \Theta$ (strikt) konvex ist.*

Satz 8.5 (Rao-Blackwell) *Sei $T : (S, \mathcal{B}) \mapsto (S', \mathcal{B}')$ suffizient für $\{Q_\vartheta : \vartheta \in \Theta\}$. Sei $h : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ ein Schätzer für $g : \Theta \mapsto \mathbb{R}$ mit $E_\vartheta|h| < \infty$ für jedes $\vartheta \in \Theta$ und $L : \Theta \times \mathbb{R} \mapsto [0, \infty)$ eine konvexe Verlustfunktion.*
Dann gilt für $h_T = E_\cdot(h|T)$

$$R(\vartheta, h_T) \leq R(\vartheta, h) \quad \forall \vartheta \in \Theta \qquad (*)$$

Ist L strikt konvex und $R(\vartheta, h) < \infty$ für jedes $\vartheta \in \mathbb{R}$, so gilt „=“ in (*) genau dann, wenn $h_T = h \quad Q_\vartheta$-f.s. $\forall \vartheta$.

Beweis: Aus der Jensen-Ungleichung[3] folgt die Behauptung. Beachte:
$R(\vartheta, h_T) = E_\vartheta(L(\vartheta, h_t)) = E_\vartheta(L(\vartheta, E_\vartheta(h|T)) \leq E_\vartheta(E_\vartheta(L(\vartheta, h)|T)) = E_\vartheta(L(\vartheta, h)) = R(\vartheta, h)$ □

Bemerkung 8.6 1. Suffizienz wird benötigt, damit $h_T = E_\vartheta(h|T)$ unabhängig von ϑ gewählt werden kann.

2. Ist h erwartungstreu, so auch h_T.

8.2 Vollständigkeit

Definition 8.7 *Sei $\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(S, \mathcal{B})$. Eine Statistik $T : (S, \mathcal{B}) \mapsto (S', \mathcal{B}')$ heißt (beschränkt)* ***vollständig*** *(für $\mathcal{Q}$), falls für jede (beschränkte) Funktion $f : (S', \mathcal{B}') \mapsto (\mathbb{R}, \mathbb{B})$ gilt:*

$$\forall \vartheta \in \Theta \quad E_\vartheta f(T) = 0 \implies \forall \vartheta \in \Theta \quad f = 0 \quad TQ_\vartheta\text{-f.s.}$$

Interpretation: Vollständigkeit ist eine Reichhaltigkeitsanforderung an $\mathcal{Q}$, also $\mathcal{Q}$ möglichst groß. Im Gegensatz dazu muss $\mathcal{Q}$ möglichst klein sein, um eine von ϑ unabhängige Version der bedingten Erwartung zuzulassen.

[3] Satz 20.6 in der Vorlesungsmitschrift zur Wahrscheinlichkeitstheorie ($\mu_\rho\varepsilon$)

Beispiel 8.8 1. $\mathcal{Q} = \{Q : Q \text{ symmetrisches Wahrscheinlichkeitsmaß auf } (\mathbb{R}, \mathbb{B})\}$[4]
Dann ist $T(x) = x$ nicht vollständig:
Wähle $f = 1_{(0,\infty)} - 1_{(-\infty,0)}$. Dann $\int f \circ T \, dQ = 0$.

2. $\mathcal{Q} = \{B^n_{(1,p)} : p \in (0,1)\}$, $T(x) = \sum_{i=1}^n x_i$
Dann ist T vollständig:
Sei $h : \{0, \ldots, n\} \mapsto \mathbb{R}$ mit $E_p h(T) = 0 \; \forall p$.
$0 = E_p h(T) = \int h \, dB_{(n,p)} = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} h(i) = \underbrace{(1-p)^n}_{>0} \quad \underbrace{\sum_{i=0}^n \binom{n}{i} h(i) \rho^i}_{\text{Polynom vom Grad } \leq n}$
mit $\rho = \frac{p}{1-p}$
$\Rightarrow \binom{n}{i} h(i) = 0 \; \forall i \Rightarrow h \equiv 0$

Definition 8.9 *1. Ein Funktional g heißt **erwartungstreu schätzbar**, falls ein erwartungstreuer Schätzer existiert.*

*2. Sei g ein Funktional und $L : \Theta \times \mathbb{R} \mapsto [0, \infty)$ eine Verlustfunktion. Ein erwartungstreuer Schätzer h^* für g heißt **gleichmäßig bester erwartungstreuer Schätzer** für g bzgl. L, falls gilt*
$\forall h$ erwartungstreuer Schätzer $\forall \vartheta \in \Theta$

$$R(\vartheta, h^*) \leq R(\vartheta, h)$$

*3. Ist $L(\vartheta, d) = (\vartheta - d)^2$, so heißt h^* in 2. **gleichmäßig bester unverfälschter minimum Varianzschätzer**.*

Satz 8.10 (Lehmann-Scheffé) *$\mathcal{Q} = \{Q_\vartheta : \vartheta \in \Theta\}$, $T : (S, \mathcal{B}) \mapsto (S', \mathcal{B}')$ suffizient und vollständig für $\vartheta \in \Theta$. Sei $h : (S, \mathcal{B}) \mapsto (\mathbb{R}, \mathbb{B})$ ein erwartungstreuer Schätzer für das Funktional $g : \Theta \mapsto \mathbb{R}$ und L eine konvexe Verlustfunktion. Dann gilt:*

1. Jeder erwartungstreuer Schätzer h^* der Form $h^* = K \circ T$ ist gleichmäßig bester erwartungstreuer Schätzer für g.

2. $h^* = E_\cdot(h|T)$ erfüllt 1.

3. Sei $L(\vartheta, \cdot)$ strikt konvex $\forall \vartheta \in \Theta$. Dann ist ein gleichmäßig bester erwartungstreuer Schätzer $\tilde{h}$ Q_ϑ-f.s. $\forall \vartheta$ bestimmt, sofern $R(\vartheta, \tilde{h}) < \infty \quad \forall \vartheta \in \Theta$.

Beweis:
1. Es gilt $E_\vartheta(h^* - E_\cdot(h|T)) = g(\vartheta) - g(\vartheta) = 0 \quad \forall \vartheta$
Wegen $h^* = K \circ T$, $E_\cdot(h|T) = \tilde{h} \circ T$ mit $\tilde{h}(s) = E_\cdot(h|T = s)$, $s \in S'$, und der Vollständigkeit von T folgt $h^* - E_\cdot(h|T) = 0 \quad Q_\vartheta$-f.s. $\forall \vartheta$ und damit nach dem Satz von Rao-Blackwell:
$R(\vartheta, h^*) = R(\vartheta, E_\cdot(h|T)) \leq R(\vartheta, h) \quad \forall \vartheta \in \Theta$
2. nach Faktorisierungslemma

[4] $\forall A \in \mathbb{B} \quad Q(A) = Q(-A)$

3. Nach dem Satz von Rao-Blackwell $R(\vartheta, \tilde{h}) = R(\vartheta, E_\cdot(\tilde{h}|T)) \quad \forall \vartheta \in \Theta$ und damit $\tilde{h} = E_\cdot(\tilde{h}|T)$ Q_ϑ-f.s. $\forall \vartheta$.
Wie in 1. impliziert die Vollständigkeit
$\tilde{h} = E_\cdot(\tilde{h}|T) = E_\cdot(h|T) \quad Q_\vartheta$-f.s. $\forall \vartheta \in \Theta$ □

Beispiel 8.11 Das Stichprobenmittel $h(x) = \frac{1}{n}\sum_{i=1}^n x_i$ ist gleichmäßig bester unverfälschter minimum Varianzschätzer für den Anteil der Raucher in der Bevölkerung, denn $T(x) = \sum_{i=1}^n x_i$ ist suffizient und vollständig für $\{B^n_{(1,p)} : p \in (0,1)\}$ (vergleiche Beispiel 7.1 und Beispiel 8.8 2.)

Satz 8.12 *Sei $\{Q_\vartheta : \vartheta \in \Theta\}$ eine k-parametrige Exponentialfamilie in $q(\vartheta) = (q_1(\vartheta), \ldots, q_k(\vartheta))$ und $T = (T_1, \ldots, T_k)$. Ist das Innere von $q(\Theta)$ nicht leer, so ist T vollständig für $\vartheta \in \Theta$.*

Beweis: Siehe Witting (1985), Satz 3.39. □

Beispiel 8.13 Für Exponentialfamilien sind folgende Statistiken T suffizient und vollständig (Sätze 7.6 und 8.12)

	Q_ϑ	Θ	T
1.	$N_{(\mu,\sigma^2)}$	$(\mu, \sigma^2) \in \mathbb{R} \times (0,\infty)$	$(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$
2.		$\mu \in \mathbb{R}$	$\sum_{i=1}^n x_i$
3.		$\sigma^2 \in (0,\infty)$	$\sum_{i=1}^n (x_i - \mu)^2$
4.	P_λ	$\lambda \in (0,\infty)$	$\sum_{i=1}^n x_i$
5.	$B_{(1,p)}$	$p \in (0,1)$	$\sum_{i=1}^n x_i$

Folgende Schätzer h sind optimal für g:
I. $E_\vartheta(id) = g(\vartheta), \quad h(x) = \frac{1}{n}\sum_{i=1}^n x_i$ optimal für 1,2,4,5
II. $Var_\vartheta(id) = g(\vartheta), \quad h(x) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$ optimal für 1
III. $Var_\vartheta(id) = g(\vartheta), \quad h(x) = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$ optimal für 3

9 Cramér-Rao-Ungleichung

Im Folgenden: $Q_\vartheta, \vartheta \in \Theta$, Wahrscheinlichkeitsmaße auf $(S, \mathcal{B})$
$L_{\vartheta_0,\vartheta} \in \frac{dQ_\vartheta}{dQ_{\vartheta_0}}$ Dichtequotienten von Q_ϑ bzgl. Q_{ϑ_0}
Falls $Q_\vartheta \ll Q_{\vartheta_0}$, dann $E_{\vartheta_0} L_{\vartheta_0,\vartheta} = 1$

Satz 9.1 (Chapman-Robbins-Ungleichung) *Sei h erwartungstreuer Schätzer für g, $\vartheta \in \Theta$ und $\Theta_0 := \{\vartheta \in \Theta \setminus \{\vartheta_0\} : Q_\vartheta \ll Q_{\vartheta_0}$ mit $L_{\vartheta_0,\vartheta} \in L_2(Q_{\vartheta_0})\}$.*
Falls $\Theta_0 \neq \emptyset$ und $Q_\vartheta \neq Q_{\vartheta_0}$ $\forall \vartheta \in \Theta_0$, dann gilt

$$Var_{\vartheta_0} h \geq \sup_{\vartheta \in \Theta_0} \frac{(g(\vartheta) - g(\vartheta_0))^2}{Var_{\vartheta_0}(L_{\vartheta_0,\vartheta})}$$

Beweis: O.E. $Var_{\vartheta_0} h < \infty$. Für $\vartheta \in \Theta_0$ gilt:
$g(\vartheta) - g(\vartheta_0) = \int h L_{\vartheta_0,\vartheta} dQ_{\vartheta_0} - \int h \; dQ_{\vartheta_0} = \int h(L_{\vartheta_0,\vartheta} - 1) dQ_{\vartheta_0} - \underbrace{\int (L_{\vartheta_0,\vartheta} - 1) dQ_{\vartheta_0}}_{0} \cdot g(\vartheta_0) =$

$E_{\vartheta_0}((h - g(\vartheta_0)) \cdot (L_{\vartheta_0,\vartheta} - 1))$
Die Cauchy-Schwarzsche Ungleichung liefert
$(g(\vartheta) - g(\vartheta_0))^2 \leq \underbrace{E_{\vartheta_0}((h - g(\vartheta_0))^2)}_{Var_{\vartheta_0} h} \cdot \underbrace{E_{\vartheta_0}((L_{\vartheta_0,\vartheta} - 1)^2)}_{Var_{\vartheta_0}(L_{\vartheta_0,\vartheta})}$ □

Die Schranke in Satz 9.1 schreiben wir als

$$\frac{\left(\frac{g(\vartheta)-g(\vartheta_0)}{\vartheta-\vartheta_0}\right)^2}{E\left(\left(\frac{L_{\vartheta_0,\vartheta}-L_{\vartheta_0,\vartheta_0}}{\vartheta-\vartheta_0}\right)^2\right)}$$

was unter geeigneten Voraussetzungen gegen die Ableitungen

$$\frac{g'(\vartheta_0)^2}{E_{\vartheta_0}\left(\left(\frac{d}{d\vartheta} L_{\vartheta_0,\vartheta}\big|_{\vartheta=\vartheta_0}\right)^2\right)}$$

konvergiert.

Im Folgenden $Q_\vartheta, \vartheta \in \Theta$, äquivalente Maße.

Definition 9.2 *Sei $\Theta \subset \mathbb{R}$ und $\vartheta_0 \in \Theta$ innerer Punkt von Θ. Dann heißt $L_{\vartheta_0,\vartheta}$* ***im quadratischen Mittel differenzierbar*** *in $\vartheta = \vartheta_0$ mit Ableitung $L'_{\vartheta_0} \in L_2(Q_{\vartheta_0})$, wenn*

$$\left(\int (L_{\vartheta_0,\vartheta} - 1 - (\vartheta - \vartheta_0) L'_{\vartheta_0})^2 dQ_{\vartheta_0}\right)^{1/2} = o(\vartheta - \vartheta_0) \qquad \text{für } \vartheta \to \vartheta_0$$

Wegen $E_{\vartheta_0}(L_{\vartheta_0,\vartheta} - 1) = 0$ und

$$E_{\vartheta_0}\left(\left|\frac{L_{\vartheta_0,\vartheta} - 1}{\vartheta - \vartheta_0} - L'_{\vartheta_0}\right|\right) \overset{\vartheta \to \vartheta_0}{\longrightarrow} 0$$

gilt $E_{\vartheta_0} L'_{\vartheta_0} = 0$.

Definition 9.3

$$I_{\vartheta_0} = E_{\vartheta_0}(L'_{\vartheta_0})^2 = Var_{\vartheta_0} L'_{\vartheta_0}$$

heißt ***Fisher-Information****.*

Satz 9.4 (Cramér-Rao-Ungleichung) *Sei $L_{\vartheta_0,\vartheta}$ im quadratischen Mittel differenzierbar in $\vartheta = \vartheta_0$. Die Fisher-Information I_{ϑ_0} sei positiv. Sei $g : \Theta \mapsto \mathbb{R}$ ein in ϑ_0 differenzierbares Funktional und h ein erwartungstreuer Schätzer für g. Dann gilt:*

$$Var_{\vartheta_0} h \geq \frac{(g'(\vartheta_0))^2}{I_{\vartheta_0}}$$

Beweis: Nach Satz 9.1 gilt $Var_{\vartheta_0} h \geq \frac{(g(\vartheta)-g(\vartheta_0))^2}{E_{\vartheta_0}(L_{\vartheta_0,\vartheta}-1)^2}$
Da $L_{\vartheta_0,\vartheta}$ im quadratischen Mittel differenzierbar ist, gilt
$E_{\vartheta_0}\left(\frac{L_{\vartheta_0,\vartheta}-1}{\vartheta-\vartheta_0}\right)^2 \to E_{\vartheta_0}(L'_{\vartheta_0})^2 = I_{\vartheta_0}$
Ebenso: $\left(\frac{g(\vartheta)-g(\vartheta_0)}{\vartheta-\vartheta_0}\right)^2 \to (g'(\vartheta_0))^2$ □

Interpretation: Die untere Schranke für die Varianz eines erwartungstreuen Schätzers ist groß, wenn das Experiment wenig Information über ϑ liefert, i.S. von I_ϑ klein.

Bemerkung 9.5 Die Cramér-Rao-Ungleichung ist i.A. schlechter als die Chapman-Robbins-Ungleichung, da nur $\vartheta = \vartheta_0$ benutzt wird.
I.A. ist die Cramér-Rao-Schranke nicht scharf. Jedoch ist sie häufig asymptotisch scharf (d.h. $n \to \infty$).

Im folgenden Beispiel ist die Cramér-Rao-Schranke scharf:

Beispiel 9.6 Sei $\{Q_\vartheta : \vartheta \in \Theta\}$ eine einparametrige, kanonische Exponentialfamilie in ϑ und T. Sei Θ offen.
Setze $g(\vartheta) = E_\vartheta T$. Dann gilt für den Schätzer $h = T$ und für jedes $\vartheta \in \Theta$
$Var_\vartheta h = g'(\vartheta) = I_\vartheta$, also $Var_\vartheta h = \frac{(g'(\vartheta))^2}{I_\vartheta}$

Beweis: h ist erwartungstreu. Q_ϑ hat μ-Dichte $c(\vartheta) \cdot e^{\vartheta T}$. Dann gilt:
$E_\vartheta e^{uT} = \int e^{uT} \cdot c(\vartheta) e^{\vartheta T} d\mu = c(\vartheta) \cdot \int e^{(u+\vartheta)T} d\mu = \frac{c(\vartheta)}{c(u+\vartheta)}$
Wir zeigen $E_\vartheta T = \frac{d}{du} E_\vartheta e^{uT}|_{u=0} = -\frac{c'(\vartheta)}{c(\vartheta)}$
Für $\delta > 0$ und $0 < |u| < \delta$ gilt
$|\frac{\exp(uT(x))-1}{u}| = |\sum_{n=0}^{\infty} \frac{u^n T(x)^n}{n!} - 1| \cdot \frac{1}{|u|} = |\sum_{n=1}^{\infty} \frac{u^{n-1} T(x)^n}{n!}| \leq \sum_{n=1}^{\infty} \frac{\delta^{n-1}|T(x)|^n}{n!} \leq \frac{\exp(\delta|T(x)|)}{\delta}$ $(*)$
Damit existiert eine integrierbare Majorante für δ hinreichend klein. Also:
$\frac{d}{du} E_\vartheta(e^{uT})|_{u=0} = E_\vartheta(\frac{d}{du} e^{uT}|_{u=0}) = E_\vartheta(T)$ und
$\frac{d}{du}\frac{c(\vartheta)}{c(u+\vartheta)}|_{u=0} = -\frac{c(\vartheta)}{c(\vartheta)^2} c'(\vartheta) = -\frac{c'(\vartheta)}{c(\vartheta)}$
Weiter ist $L_{\vartheta_0,\vartheta} = \frac{c(\vartheta)e^{\vartheta T}}{c(\vartheta_0)e^{\vartheta_0 T}}$ und damit
$L'_{\vartheta_0} = \frac{c'(\vartheta_0)}{c(\vartheta_0)} + T = T - E_{\vartheta_0} T$ (Wie in $(*)$ zeigt man, dass eine integrierbare Majorante von $\left(\frac{L_{\vartheta_0,\vartheta}-1}{\vartheta-\vartheta_0}\right)^2$ existiert.)
Also $I_\vartheta = Var_\vartheta L'_\vartheta = Var_\vartheta T$
Bleibt zu zeigen: $g'(\vartheta) = Var_\vartheta T$
$g'(\vartheta) = \frac{d}{d\vartheta} E_\vartheta T = \frac{d}{d\vartheta}\frac{\int T e^{\vartheta T} d\mu}{\int e^{\vartheta T} d\mu} = \frac{\int T^2 e^{\vartheta T} d\mu}{\int e^{\vartheta T} d\mu} - \frac{\int T e^{\vartheta T} d\mu}{(\int e^{\vartheta T} d\mu)^2} \int T e^{\vartheta T} d\mu = E_\vartheta T^2 - (E_\vartheta T)^2 = Var_\vartheta T$

(Die Funktion $\vartheta \mapsto \int e^{\vartheta T} d\mu$ ist ∞ oft differenzierbar mit $\frac{d^n}{d\vartheta^n} \int e^{\vartheta T} d\mu = \int T^n e^{\vartheta T} d\mu$; zum Beweis verwende man $(*)$). □

Bemerkung 9.7 Offenbar nimmt T die Cramér-Rao-Schranke an, ist also gleichmäßig bester unverfälschter minimum Varianzschätzer für $g(\vartheta) = E_\vartheta T$. Das folgt auch schon aus Satz 8.13 und Satz 8.10 (Lehmann-Scheffé).

Kapitel III

Regressionsanalyse

10 Lineare Regression mit festem Design

10.1 Allgemeine Regression

$M \subset \mathbb{R}^k$ Messstellen, Zeitpunkte etc.
$Q(\cdot|x)$ Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathbb{B})$, $x \in M$
Gegeben seien (bzw. wähle) $x_1, \ldots, x_n \in M$ und Realisationen $y_1, \ldots, y_n$ mit y_i Realisation nach $Q(\cdot|x_i)$, $i = 1, \ldots, n$.

Aufgabe: Schätze Regressionsfunktion m mit $m(x) = \int zQ(dz|x)$, $x \in M$, basierend auf $(x_1, y_1), \ldots, (x_n, y_n)$ (unter gewissen Glattheitsannahmen an m).
(Andere Funktionale:
$m_1(x)$ Median von $Q(\cdot|x)$
$m_2(x) = \int h \; dQ(\cdot|x)$)

Nun zu einer Spezialisierung der allgemeinen Regression:

10.2 Einfache lineare Regression

Die Regressionsfunktion m ist eine Gerade.
$m(x) = \int zQ(dz|x) = \alpha x + \beta$ **Regressionsgerade**
α *Regressionssteigung*
β *Achsenabschnitt*

Beispiel 10.1 Für die Ausdehnung $m(x)$ einer Federwaage in Abhängigkeit von der Kraft x gilt das physikalische Gesetz $m(x) = \alpha x + \beta$, $x \in M = [0, M_{max}]$. Eine Messung mit Gewichten $x_1, \ldots, x_n \in M$ führt zu Auslenkungen $y_1, \ldots, y_n$.
In der Regel werden $(x_1, y_1), \ldots, (x_n, y_n)$ nicht genau auf einer Geraden liegen (Messfehler!).

Aufgabe: Basierend auf $(x_1, y_1), \ldots, (x_n, y_n)$ schätze Regressionsgerade $\hat{m}(x) = \hat{\alpha} x + \hat{\beta}$.

Kleinste-Quadrate-Schätzer:
Wähle $(\hat{\alpha}, \hat{\beta})$ so, dass
$L(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha x_i - \beta)^2 = \min_{\alpha,\beta}!$
$\frac{\partial L(\alpha,\beta)}{\partial \alpha} = 0 \Leftrightarrow \sum_{i=1}^n (y_i - \alpha x_i - \beta) x_i = 0$ $(*)$
$\frac{\partial L(\alpha,\beta)}{\partial \beta} = 0 \Leftrightarrow \sum_{i=1}^n (y_i - \alpha x_i - \beta) = 0$ $(**)$
Aus $(**)$ folgt:

$$\hat{\beta} = \overline{y} - \hat{\alpha}\overline{x}$$

Damit und mit $(*)$ folgt
$\sum_{i=1}^n (y_i - \overline{y} - \alpha(x_i - \overline{x}))x_i = 0 \Leftrightarrow \hat{\alpha} = \frac{\sum_{i=1}^n (y_i - \overline{y})x_i}{\sum_{i=1}^n (x_i - \overline{x})x_i} \Leftrightarrow$

$$\hat{\alpha} = \frac{\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n y_i (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

Bemerkung 10.2 $(\hat{\alpha}, \hat{\beta})$ heißt **Kleinste-Quadrate-Schätzer**.

Formulierung mit Zufallsvariablen:
Gegeben $x_1, \ldots, x_n$ unabhängige Variablen (Messstellen),
$y_1, \ldots, y_n$ abhängige Variablen (Messergebnisse).
Betrachte y_i als Realisation einer Zufallsvariablen Y_i mit $\mathcal{L}(Y_i) = Q(\cdot|x_i)$
Lineare Regressionsannahme:
$EY_i = \int z \, d\mathcal{L}(Y_i)(z) = m(x_i) = \alpha x_i + \beta$
Also

$$Y_i = \alpha x_i + \beta + \varepsilon_i \text{ mit } E\varepsilon_i = 0 \tag{10.1}$$

$\varepsilon_i (= Y_i - \alpha x_i - \beta)$ heißen **Residuen**.

Im Folgenden $\hat{\beta} = \overline{Y} - \hat{\alpha}\overline{x}, \quad \hat{\alpha} = \frac{\sum_{i=1}^n (Y_i - \overline{Y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$

Lemma 10.3 (Erwartungstreue) *Es gelte (10.1). Dann*

1. $E\hat{\alpha} = \alpha$
2. $E\hat{\beta} = \beta$

Beweis: $E\hat{\alpha} = \frac{\sum_{i=1}^n (EY_i - E\overline{Y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (\alpha x_i + \beta - \frac{1}{n}\sum_{j=1}^n (\alpha x_j + \beta))(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})} = \frac{\sum_{i=1}^n \alpha (x_i - \overline{x})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \alpha$
$E\hat{\beta} = E\overline{Y} - \overline{x}E\hat{\alpha} = \alpha\overline{x} + \beta - \alpha\overline{x} = \beta$ □

10.3 Multiple lineare Regression

$x = (x_1, \ldots, x_{k-1}) \in M \subset \mathbb{R}^{k-1}$
$Q(\cdot|x)$ Wahrscheinlichkeitmaß auf $(\mathbb{R}, \mathbb{B})$, $x \in M$
$m(x) = \int z Q(dz|x) = \beta_0 + \sum_{i=1}^{k-1} \beta_i x_i$

Jetzt $x_i = (x_{i,1}, \ldots, x_{i,k-1}) \in M$
Y_i Zufallsvariable mit $\mathcal{L}(Y_i) = Q(\cdot\,|x_i)$
Modell:

$$Y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j x_{i,j} + \varepsilon_i \text{ mit } E\varepsilon_i = 0,\ i = 1, \ldots, n \tag{10.2}$$

Speziell: Polynomiale Regression $x_{i,j} = x_{i,1}^j$

Matrix-Schreibweise:
$Y = (Y_1, \ldots, Y_n)^\top, \quad y = (y_1, \ldots, y_n)^\top, \quad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top, \quad \beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})^\top$

Bezeichnung 10.4 *Die $n \times k$-Matrix*

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k-1} \end{pmatrix}$$

heißt ***Design-Matrix****.*

Bezeichnung 10.5 *1.* $E(Y) = (E(Y_1), \ldots, E(Y_n))^\top$

2. $E\varepsilon = 0$*, falls* $E\varepsilon_i = 0,\ i = 1, \ldots, n$

Damit

$$Y = X\beta + \varepsilon \text{ mit } E\varepsilon = 0 \tag{10.3}$$

Im Folgenden sei der Stichprobenumfang $n \geq k$.

Kleinste-Quadrate-Schätzer:
Bestimme $\hat{\beta}$ so dass
$L(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^{k-1} \beta_j x_{ij})^2 = ||y - X\beta||^2 = \min_\beta$ $(*)$
wobei $||\cdot||$ den euklidischen Abstand im $\mathbb{R}^n$ bezeichnet.

Setze $L_X = \{X\beta : \beta \in \mathbb{R}^k\} = \text{span}\left(\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x_{1,k-1} \\ \vdots \\ x_{n,k-1} \end{pmatrix}\right)$.
Dann ist L_X ein linearer Unterraum des $\mathbb{R}^n$.

Lemma 10.6 *1. (Normalengleichung) Hinreichende und notwendige Bedingung für Lösungen von* $(*)$ *ist*

$$X^\top X\beta = X^\top y$$

und es existiert mindestens eine Lösung von $(*)$*.*

2. *Falls $X^\top X$ invertierbar ist, dann ist*

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

eindeutige Lösung von $(*)$.

Beweis:

1. Sei $\tilde{\beta} \in \mathbb{R}^k$. Dann gilt
   $||y - X\tilde{\beta}|| \leq ||y - X\beta|| \quad \forall \beta \in \mathbb{R}^k$
   $\Leftrightarrow X\tilde{\beta}$ orthogonale Projektion von y in L_X
   $\Leftrightarrow < X\beta, y - X\tilde{\beta} >= 0 \quad \forall \beta$
   $\Leftrightarrow (X\beta)^\top y = (X\beta)^\top X\tilde{\beta} \quad \forall \beta$
   $\Leftrightarrow \beta^\top X^\top y = \beta^\top X^\top X\tilde{\beta} \quad \forall \beta$
   $\Leftrightarrow X^\top y = X^\top X\tilde{\beta}$
   Die Existenz einer Lösung von $(*)$ folgt aus der Existenz der orthogonalen Projektion.
2. klar □

Bemerkung 10.7 Falls $n \geq k$, dann

$$X^\top X \text{ invertierbar} \Leftrightarrow \text{Rang}(X) = k \Leftrightarrow \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{n,j} \end{pmatrix}, \quad j = 1, \ldots, k-1 \text{ linear unabhängig}$$

Im Folgenden $\text{Rang}(X) = k$ (sonst keine eindeutigen Kleinste-Quadrate-Schätzer).

Lemma 10.8 *Seien $Y_1, \ldots, Y_n$ Zufallsvariablen, $c \in \mathbb{R}^m$, A $m \times n$-Matrix und $Z := c + AY$.*

1. *Falls $E|Y_i| < \infty, \quad i = 1, \ldots, n$, dann*

$$EZ = c + AEY$$

2. *Falls $EY_i^2 < \infty, \quad i = 1, \ldots, n$, dann*

$$\Sigma_Z := (Cov(Z_i, Z_j))_{i,j} = A \cdot \Sigma_Y \cdot A^\top$$

Beweis:

1. $EZ_j = E(c_j + \sum_{i=1}^n a_{ji} Y_i) = c_j + \sum_{i=1}^n a_{ji} EY_i$

2. $Cov(Z_i, Z_j) = E((Z_i - EZ_i)(Z_j - EZ_j)) = E((\sum_{r=1}^n a_{ir}(Y_r - EY_r))(\sum_{l=1}^n a_{jl}(Y_l - EY_l))) =$
   $\sum_{r=1}^n a_{ir} \sum_{l=1}^n a_{jl} Cov(Y_r, Y_l) = (a_{i1}, \ldots, a_{in}) \Sigma_Y \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jn} \end{pmatrix}$ (i-te Zeile von A bzw. j-te Spalte von $A^\top$) □

$\hat{\beta} = (X^\top X)^{-1} X^\top Y$ Kleinste-Quadrate-Schätzer.

Lemma 10.9 (Erwartungstreue) *Unter den Bedingungen (10.3) gilt*

$$E\hat{\beta} = \beta$$

Beweis: Wegen $Y = X\beta + \varepsilon$ gilt:
$\hat{\beta} = (X^\top X)^{-1} X^\top (X\beta + \varepsilon) = \underbrace{(X^\top X)^{-1} X^\top X \beta}_{\beta} + (X^\top X)^{-1} X^\top \varepsilon$

Nach Lemma 10.8 1. $E\hat{\beta} = \beta + (X^\top X)^{-1} X^\top E\varepsilon = \beta$ □

Lemma 10.10 (Kovarianzmatrix) *Die Residuen ε_i seien paarweise unkorreliert und $Var(\varepsilon_i) = \sigma^2, \quad i = 1, \ldots, n$. Dann gilt*

$$\Sigma_{\hat{\beta}} = \sigma^2 (X^\top X)^{-1}$$

Beweis: Nach dem Beweis von Lemma 10.9 gilt $\hat{\beta} = \beta + (X^\top X)^{-1} X^\top \varepsilon$ und damit
$\Sigma_{\hat{\beta}} = (X^\top X)^{-1} X^\top \Sigma_\varepsilon ((X^\top X)^{-1} X^\top)^\top = (X^\top X)^{-1} X^\top (\sigma^2 I_n) X ((X^\top X)^{-1})^\top = \sigma^2 ((X^\top X)^\top)^{-1} = \sigma^2 (X^\top X)^{-1}$, wobei I_n die Einheitsmatrix ist. □

Bezeichnung 10.11 *1. $\tilde{\beta}$ heißt **erwartungstreuer Schätzer** für β, falls $E\tilde{\beta} = \beta$*

*2. $\tilde{\beta}$ heißt **linearer Schätzer** für β, falls eine Matrix A existiert mit $\tilde{\beta} = AY$*

Offenbar ist der Kleinste-Quadrate-Schätzer ein linearer erwartungstreuer Schätzer.

Standardmodell: (10.4)

$$Y = X\beta + \varepsilon$$
$$E\varepsilon = 0, \quad \varepsilon_i \text{ unkorreliert}, \quad Var(\varepsilon_i) = \sigma^2, \quad i = 1 \ldots, n$$
$$n \geq k, \quad \text{Rang}(X) = k$$

Bezeichnung 10.12 *Eine Matrix A heißt positiv semidefinit, falls*

$$\forall v: \quad v^\top A v \geq 0$$

Bemerkung 10.13 Sei C eine Matrix. Dann ist $C^\top C$ postitiv semidefinit, denn
$v^\top C^\top C v = (Cv)^\top Cv = ||Cv||^2 \geq 0$.

Satz 10.14 (Gauß-Markov) *Im Modell (10.4) gilt für jeden linearen, erwartungstreuen Schätzer $\tilde{\beta} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_{k-1})$:*

$$Var(\hat{\beta}_i) \leq Var(\tilde{\beta}_i) \qquad i = 0, \ldots, k-1$$

Beweis: Sei $\tilde{\beta} = AY$ mit $E\tilde{\beta} = \beta$. Es gilt $\beta = E\tilde{\beta} = AEY = AE(X\beta + \varepsilon) = AX\beta$, da $E\varepsilon = 0$.
Also wegen β beliebig $AX = I_k$ (1)
Wir zeigen: $Var(v^\top \hat{\beta}) \leq Var(v^\top \tilde{\beta}) \quad \forall v \in \mathbb{R}^k$ (2)
Hieraus folgt die Behauptung mit der Wahl von v als Einheitsvektor.
Zu (2):
$Var(v^\top \hat{\beta}) \overset{10.8}{=} \overset{2.}{} v^\top \Sigma_{\hat{\beta}} v \overset{10.10}{=} \sigma^2 v^\top (X^\top X)^{-1} v$ und
$Var(v^\top \tilde{\beta}) = Var(v^\top AY) \overset{10.8}{=} \overset{2.}{} v^\top A \Sigma_Y (v^\top A)^\top = v^\top A \Sigma_\varepsilon A^\top v = \sigma^2 v^\top A A^\top v$
Es gilt (2), falls $v^\top (AA^\top - (X^\top X)^{-1}) v \geq 0$, d.h. falls $AA^\top - (X^\top X)^{-1}$ positiv semidefinit ist. Dies folgt aus der Darstellung
$AA^\top - (X^\top X)^{-1} = (A - (X^\top X)^{-1} X^\top)(A - (X^\top X)^{-1} X^\top)^\top$ (3)
und Bemerkung 10.13.
Zu (3):
$(A - (X^\top X)^{-1} X^\top)(A^\top - X((X^\top X)^{-1})^\top) = AA^\top - (X^\top X)^{-1} X^\top A^\top - AX(X^\top X)^{-1} +$
$(X^\top X)^{-1} = AA^\top - (X^\top X)^{-1}$, da $AX = I_k, \quad X^\top A^\top = (AX)^\top = I_k$ □

11 Regression mit zufälligem Design

Bisher waren $x_1, \ldots, x_n$ fest. Nun gehen wir von Zufallsvariablen $X_1, \ldots, X_n$ aus.

Seien $(X_1, Y_1), \ldots, (X_n, Y_n) : (\Omega, \mathcal{A}, P) \mapsto (\mathbb{R}^{k-1} \times \mathbb{R}, \mathbb{B}^{k-1} \times \mathbb{B})$ i.i.d. Zufallsvariablen mit $E|Y_i| < \infty$.
Setze $Q(\cdot\,|x) = P(Y_1 \in \cdot\,|X_1 = x)$ bedingte Verteilung von Y_1 gegeben $X_1 = x$.

Aufgabe: Schätze Regressionsfunktion m mit

$$m(x) = \int z Q(dz|x) = E(Y_1|X_1 = x)$$

basierend auf i.i.d. Zufallsvariablen (X_i, Y_i).

11.1 Lineare Regression

Im Folgenden: $m(x) = \beta_0 + \sum_{i=1}^{k-1} \beta_i x_i, \quad x \in \mathbb{R}^{k-1}$

Matrix-Schreibweise:

$$X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk-1} \end{pmatrix}$$ **„zufällige" Design-Matrix**

wobei $X_i = (X_{i1}, \ldots, X_{ik-1}) : (\Omega, \mathcal{A}, P) \mapsto (\mathbb{R}^{k-1}, \mathbb{B}^{k-1})$.
(Zur Messbarkeit von X: Fasse $n \times k$-Matrix als Element des $\mathbb{R}^{n \cdot k}$ auf. Dann folgt die Messbarkeit von X aus der Messbarkeit von $(1, X_{11}, \ldots, X_{1k-1}, \ldots, 1, X_{n1}, \ldots, X_{nk-1})$.)

Kleinste-Quadrate-Schätzer im zufälligen Design:

Analog zu Lemma 10.6 2.:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y$$

falls $X^\top X$ invertierbar.

Speziell einfache lineare Regression:
$m(x) = \alpha x + \beta$
$\hat{\beta} = \overline{Y} - \hat{\alpha}\overline{X}$ mit $\overline{X} = \frac{1}{n}\sum X_i$

$\hat{\alpha} = \frac{\sum_{i=1}^n (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2}$ falls $\sum_{i=1}^n (X_i - \overline{X})^2 > 0$, also $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ linear unabhängig bzw. nach Bemerkung 10.7 $X^\top X$ invertierbar.

Lemma 11.1 *Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ unabhängige Zufallsvariablen, $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$. Dann gilt:*

$$P(Y \in \cdot \,| X = x) = \bigotimes_{i=1}^n P(Y_i \in \cdot \,| X_i = x_i)$$

Beweis: Siehe Übung. □

Lemma 11.2 *Sei $X^\top X$ P-f.s. invertierbar, $\hat{\beta} := (X^\top X)^{-1} X^\top Y$. Dann*

$$E\hat{\beta} = \beta$$

Beweis: Setze $g((X_1, \ldots, X_n), Y) = (X^\top X)^{-1} X^\top Y = \hat{\beta}$.
Nach Lemma 6.5 gilt:
$E(\hat{\beta}_i | (X_1, \ldots, X_n) = x) = E(g_i((X_1, \ldots, X_n), Y) | (X_1, \ldots, X_n) = x)$
$= \int g_i(x, y) P(Y \in dy | (X_1, \ldots, X_n) = x) \overset{11.1}{=} \int g_i(x, y) d\mathcal{L}(\tilde{Y})(y) = E(g_i(x, \tilde{Y}))$
wobei $\tilde{Y} = (\tilde{Y}_1, \ldots \tilde{Y}_n)$ Zufallsvariable mit $\mathcal{L}(\tilde{Y}) = \bigotimes_{i=1}^n P(Y_i \in \cdot \,| X_i = x_i)$
$(E\tilde{Y}_i = E(Y_i | X_i = x) = m(x))$
Aus Lemma 10.9 folgt $E(g_i(x, \tilde{Y})) = \beta_i$
Also gilt $E\hat{\beta}_i = \int \underbrace{E(\hat{\beta}_i | (X_1, \ldots, X_n) = x)}_{\beta_i \ \mathcal{L}(X_1, \ldots, X_n)\text{-f.s.}} d\mathcal{L}(X_1, \ldots, X_n)(x) = \beta_i$ □

11.2 Nichtparametrische Regression

Sei $(X, Y) : (\Omega, \mathcal{A}, P) \mapsto (\mathbb{R}^2, \mathbb{B}^2)$ mit $E|Y| < \infty$.

Aufgabe: Schätze Regressionsfunktion m mit

$$m(x) = E(Y | X = x)$$

basierend auf i.i.d. Zufallsvariablen $(X_1, Y_1), \ldots, (X_n, Y_n)$ mit $\mathcal{L}(X_i, Y_i) = \mathcal{L}(X, Y)$.

Beachte: $m(x) = \int z \; dF(\cdot|x)(z)$ mit $F(y|x) = P(Y \leq y | X = x)$ bedingte Verteilungsfunktion[1].

Falls $P\{X = x\} > 0$:

$$\frac{P\{Y \leq y, X = x\}}{P\{X = x\}} = P(Y \leq y | X = x)$$

Plausibel ist

$$\frac{P\{Y \leq y, X \in [x - a_n, x + a_n]\}}{P\{X \in [x - a_n, x + a_n]\}} \to P(Y \leq y | X = x)$$

falls $a_n \downarrow 0, \quad P\{X \in [x - a_n, x + a_n]\} > 0.$

Zu Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ setze

$$F_n(y|x) := \frac{\frac{1}{n} \sum_{i=1}^n 1_{[x-a_n, x+a_n] \times (-\infty, y]}(x_i, y_i)}{\frac{1}{n} \sum_{i=1}^n 1_{[x-a_n, x+a_n]}(x_i)}$$

Bezeichnung 11.3 *1. $F_n(\cdot|x)$ heißt **empirische bedingte Verteilungsfunktion**.*

*2. $Q_n(\cdot|x) = \int dF_n(\cdot|x) = \frac{\sum_{i=1}^n 1_{[x-a_n, x+a_n]}(x_i) \varepsilon_{y_i}}{\sum_{i=1}^n 1_{[x-a_n, x+a_n]}(x_i)}$ heißt **empirische bedingte Verteilung**.*

Plausibel ist[2]:

$$m_n(x) = \int z \; dF_n(\cdot|x)(z) = \frac{\sum_{i=1}^n y_i 1_{[x-a_n, x+a_n]}(x_i)}{\sum_{i=1}^n 1_{[x-a_n, x+a_n]}(x_i)}$$

ist eine nichtparametrische Schätzung für $m(x) = E(Y|X = x)$.

12 Lineare Modelle mit Normalverteilungsannahme

Im Folgenden: X $n \times k$-Design-Matrix, $\text{Rang}(X) = k$, festes Design

Standardmodell (10.4):
$Y = X\beta + \varepsilon$ mit $\beta \in \mathbb{R}^k$, $n \geq k$, ε_i unkorreliert, $E\varepsilon_i = 0$, $Var\varepsilon_i = \sigma^2$

Neue Darstellung (12.5)

$$Y = \mu + \sigma W$$

Lineares Modell mit $\mu \in L_k = \{X\beta : \beta \in \mathbb{R}^k\}$ linearer Teilraum des $\mathbb{R}^n$, $W = (W_1, \ldots, W_n)^\top$, W_i unkorreliert, $EW_i = 0$, $VarW_i = 1$

[1] oder Verteilungsfunktion der bedingten Verteilung

[2] Gleitendes Mittel

Für die Kleinste-Quadrate-Schätzer in (10.4) bzw. (12.5) gilt gemäß Satz 10.6
$\hat{\mu} = X\hat{\beta} = \text{Proj}(Y, L_k)$ orthogonale Projektion von Y in L_k.

Häufig ist der Übergang zu einer kanonischen Form des linearen Modells sinnvoll:

Kanonische Form (12.6)

Es existiert eine orthonormale Basis $e_1, \ldots, e_n$ des $\mathbb{R}^n$ mit $L_k = \text{span}(e_1, \ldots, e_k)$. Dann bildet $\mathcal{T} = (e_1, \ldots, e_n)^\top$ eine orthogonale Matrix mit
$\mathcal{T}e_j = (< e_1, e_j >, \ldots, < e_n, e_j >)^\top = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ (1 an der j-ten Stelle) und $\mathcal{T}L_k = \mathbb{R}^k \times \{0\}^{n-k}$
Dann: $\mathcal{T}Y = \mathcal{T}\mu + \sigma\mathcal{T}W$ bzw.

$$\tilde{Y} = \nu + \sigma\tilde{W} \qquad (*)$$

Offenbar bildet $(*)$ ein neues lineares Modell, wobei $\nu \in \mathcal{T}L_k = \mathbb{R}^k \times \{0\}^{n-k}$.

Für den Kleinste-Quadrate-Schätzer $\hat{\nu}$ gilt
$\hat{\nu} \in \mathbb{R}^k \times \{0\}^{n-k}$ und $\hat{\nu} = (\tilde{Y}_1, \ldots, \tilde{Y}_k, 0, \ldots, 0)^\top$
Mit $\mathcal{T}^\top = \mathcal{T}^{-1}$ und $||\tilde{Y} - \hat{\nu}|| = ||\mathcal{T}^\top\tilde{Y} - \mathcal{T}^\top\hat{\nu}|| = ||Y - \underbrace{\mathcal{T}^\top\hat{\nu}}_{\in L_k}||$ folgt $\hat{\mu} = \mathcal{T}^\top\hat{\nu}$ oder

$$\mathcal{T}\hat{\mu} = \hat{\nu} \qquad (**)$$

Bezeichnung 12.1 *Sei $\hat{\mu} = Proj(Y, L_k)$. Dann heißt*

$$\hat{\sigma}^2 = \frac{1}{n-k}||Y - \hat{\mu}||^2$$

Residualschätzer.

Bemerkung 12.2 Es ist $\hat{\sigma}^2 = \frac{1}{n-k}||Y-\hat{\mu}||^2 = \frac{1}{n-k}||\mathcal{T}Y-\mathcal{T}\hat{\mu}||^2 = \frac{1}{n-k}||\tilde{Y}-\hat{\nu}||^2 = \frac{1}{n-k}\sum_{i=k-1}^{n}\tilde{Y}_i^2$

Satz 12.3 *Sei $EW_i = 0$, $VarW_i = 1$, $i = 1, \ldots, n$. Dann gilt*

1. *$\hat{\sigma}^2$ ist erwartungstreuer Schätzer für σ^2.*
2. *Sind $W_1, \ldots, W_n$ stochastisch unabhängig, $EW_i^3 = 0$, $EW_i^4 < \infty$, so sind $\hat{\mu}$ und $\hat{\sigma}^2$ unkorreliert.*
3. *Sind $W_1 \ldots, W_n$ i.i.d. $N_{(0,1)}$-verteilt, so sind $\hat{\mu}$ und $\hat{\sigma}^2$ stochastisch unabhängig und $\frac{n-k}{\sigma^2}\hat{\sigma}^2$ ist $\mathcal{X}^2_{n-k}$-verteilt.*

Beweis:
1. Wir verwenden das kanonische Modell und erhalten aus $(**)$
   $Y - \hat{\mu} = \mathcal{T}^\top(\tilde{Y} - \hat{\nu}) = \mathcal{T}^\top(0, \ldots, 0, \tilde{Y}_{k+1}, \ldots, \tilde{Y}_n)^\top \Rightarrow ||Y - \hat{\mu}||^2 = ||\tilde{Y} - \hat{\nu}||^2 = \sum_{i=k+1}^{n}\tilde{Y}_i^2$
   Wegen $E\tilde{Y}_i^2 = E\sigma^2\tilde{W}_i^2 = \sigma^2$, $i = k+1, \ldots, n$ folgt 1.

2. Wir zeigen: $E((\hat{\mu}-\mu)\hat{\sigma}^2)=0$
   Sei $Z=(Z_1,\ldots,Z_n)^\top = Y-\mu=\sigma W$ und P Projektionsmatrix in L_k, also $\hat{\mu}=PY$, $P\mu=\mu \quad \mu\in L_k$
   Dann $\hat{\mu}-\mu=P(Y-\mu)=PZ$ und
   $(n-k)\hat{\sigma}^2=||Y-\hat{\mu}||^2=||(I_n-P)Y||^2=Y^\top(I_n-P)^\top(I_n-P)Y=(Z+\mu)^\top(I_n-P)^\top(I_n-P)(Z+\mu)\overset{(I_n-P)\mu=0}{=}Z^\top(I_n-P)^\top(I_n-P)Z$
   Damit: $(n-k)E((\hat{\mu}-\mu)\hat{\sigma}^2)=E(PZZ^\top(I_n-P)^\top(I_n-P)Z)=P(\sum_{i,j}c_{ij}E(Z_lZ_iZ_j))_{l=1,\ldots,n}$ mit geeigneten $c_{ij}\in\mathbb{R}$.
   Wegen $E|Z_1Z_2Z_3|\le E(Z_1^2Z_2^2)+\underbrace{EZ_3^2}_{<\infty}$, $E(Z_1^2Z_2^2)\le EZ_1^4+EZ_2^4<\infty$ und $E|Z_1^2Z_2|\le EZ_1^4+EZ_2^2<\infty$ gilt $E|Z_lZ_jZ_i|<\infty \quad \forall i,j,l$
   Aus der Unabhängigkeit folgt
   $E(Z_lZ_jZ_i)=0\ \forall i,j,l\in\{1,\ldots,n\}$
3. Sind $W_1,\ldots,W_n$ i.i.d. $N_{(0,1)}$-verteilt und $\tilde{W}=\mathcal{T}W=\mathcal{T}(W_1,\ldots,W_n)^\top$, so folgt aus Satz 4.3: $\tilde{W}_1,\ldots,\tilde{W}_n$ i.i.d. $N_{(0,1)}$-verteilt. Damit $\tilde{Y}_1,\ldots,\tilde{Y}_n$ stochastisch unabhängig.
   Wegen $\hat{\mu}=\mathcal{T}^\top\hat{\nu}=\mathcal{T}^\top(\tilde{Y}_1,\ldots,\tilde{Y}_k,0,\ldots,0)$ und $\hat{\sigma}^2=\frac{1}{n-k}\sum_{i=k+1}^n\tilde{Y}_i^2$ folgt die stochastische Unabhängigkeit von $\hat{\mu}$ und $\hat{\sigma}^2$ und es gilt
   $\frac{n-k}{\sigma^2}\hat{\sigma}^2=\frac{1}{\sigma^2}\sum_{i=k+1}^n\sigma^2\tilde{W}_i^2 \qquad \mathcal{X}^2_{n-k}$-verteilt.

Bezeichnung 12.4 *Falls $W_1,\ldots,W_n$ i.i.d. $N_{(0,1)}$-verteilt, so heißt (12.5)* ***lineares Modell mit Normalverteilungsannahme (LMN).***

Bemerkung 12.5 Gemäß Satz 4.3 ist das kanonische Modell (12.6) LMN, falls das Modell (12.5) LMN ist. Es gilt also $\tilde{W}_1,\ldots,\tilde{W}_n$ i.i.d. $N_{(0,1)}$-verteilt.

Satz 12.6 *Es liege ein LMN vor. Sei $g: L_k\mapsto\mathbb{R}$ eine lineare Abbildung. Dann sind $\hat{\sigma}^2$ und $g(\hat{\mu})$ gleichmäßig beste unverfälschte minimum Varianzschätzer für σ^2 und g.*

Beweis: Mit $\hat{\mu}=X\hat{\beta}$, $\mu=X\beta$ und Lemma 10.9 folgt
$Eg(\hat{\mu})=Eg(X\hat{\beta})=g(X\underbrace{E\hat{\beta}}_{\beta})=g(\mu)$
Satz 12.3 liefert $E\hat{\sigma}^2=\sigma^2$.
Durch Übergang zum kanonischen Modell $\tilde{Y}=\mathcal{T}Y=\mathcal{T}\mu+\sigma\mathcal{T}W=\nu+\sigma\tilde{W}$ erhalten wir für die Verteilung von $\tilde{Y}$ unter $\nu=(\nu_1,\ldots,\nu_k,0,\ldots,0)$ und σ^2 die Dichte (bzgl. λ^n)
$\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\nu_i)^2\right)$
$=\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n y_i^2-2\sum_{i=1}^k\nu_iy_i+\sum_{i=1}^k\nu_i^2\right)\right),\quad y\in\mathbb{R}^n$
Offenbar liegt eine k-parametrige Exponentialfamilie mit der vollständigen (Satz 8.12) und suffizienten (Satz 7.6) Statistik $T(y)=(\sum_{i=1}^n y_i^2,y_1,\ldots,y_k)$ und $q(\eta)=-\frac{1}{2\eta_1}(1,-2\eta_2,\ldots,-2\eta_{k+1})$ mit $\eta=(\eta_1,\ldots,\eta_{k+1})=(\sigma^2,\nu_1,\ldots,\nu_k)\in(0,\infty)\times\mathbb{R}^k$ vor.
Da $\hat{\sigma}^2$ und $g(\hat{\mu})$ erwartungstreue Schätzer sind, die nur von T abhängen (vgl. Bezeichnung 12.1 und $(**)$), folgt die Optimalität von $\hat{\sigma}^2$ und $\hat{\mu}$ aus dem Satz von Lehmann-Scheffé (Satz 8.10).□

Beispiel 12.7 (k-Stichprobenproblem) Man vergleiche k Behandlungen, wobei die i-te n_j-mal unabhängig wiederholt wird und die Streuung in allen Versuchen gleich ist.
Sei $n = \sum_{i=1}^{k} n_i$. Wir wählen das lineare Modell
$Y_i = \beta_1 + \sigma W_i$ falls $1 \le i \le n_1$
$Y_i = \beta_2 + \sigma W_i$ falls $n_1 \le i \le n_1 + n_2$
$\vdots$
$Y_i = \beta_k + \sigma W_i$ falls $n_1 + \ldots + n_{k-1} \le i \le n$
Dabei sind $\sigma, \beta_1, \ldots, \beta_k$ unbekannt. Sei

$$X = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \cdot \\ 1 & 0 & & \cdot \\ 0 & 1 & & \cdot \\ \cdot & \vdots & & \cdot \\ \cdot & 1 & & \cdot \\ \cdot & 0 & \ddots & 0 \\ \cdot & \cdot & & 1 \\ \cdot & \cdot & & \vdots \\ 0 & 0 & & 1 \end{pmatrix}$$

Dann $Y = X\beta + \sigma W, \quad \beta \in \mathbb{R}^k, \quad \sigma > 0$
Es gilt:

$$X^\top X = \begin{pmatrix} n_1 & & 0 \\ & \ddots & \\ 0 & & n_k \end{pmatrix}, \quad (X^\top X)^{-1} = \begin{pmatrix} n_1^{-1} & & 0 \\ & \ddots & \\ 0 & & n_k^{-1} \end{pmatrix}$$

$X^\top Y = (\sum_{i=1}^{n_1} Y_i, \sum_{i=n_1+1}^{n_1+n_2} Y_i, \ldots, \sum_{i=n_1+\ldots+n_{k-1}+1}^{n} Y_i)^\top =: (n_1 \overline{Y}_{(1)}, \ldots, n_k \overline{Y}_{(k)})^\top$
Damit
$\hat{\beta} = (X^\top X)^{-1} X^\top Y = (\overline{Y}_{(1)}, \ldots, \overline{Y}_{(k)})^\top$
$\hat{\mu} = X\hat{\beta} = (\underbrace{\overline{Y}_{(1)}, \ldots, \overline{Y}_{(1)}}_{n_1}, \overline{Y}_{(2)}, \ldots, \underbrace{\overline{Y}_{(k)}, \ldots, \overline{Y}_{(k)}}_{n_k})^\top$
Der Residualschätzer berechnet sich zu:
$\sigma^2 = \frac{1}{n-k} ||Y - \hat{\mu}||^2 = \frac{1}{n-k} \sum_{j=1}^{k} \sum_{i=n_0+\ldots+n_{j-1}+1}^{n_1+\ldots+n_j} (Y_i - \overline{Y}_{(j)})^2, \quad n_0 = 0$

12.1 Tests bei linearen Hypothesen

Im Folgenden liege ein lineares Modell $Y = \mu + \sigma W$ mit Normalverteilungsannahme vor, also $W_1, \ldots, W_n$ i.i.d. $N_{(0,1)}$-verteilt.
Sei $L_h \subset L_k$ ein weiterer h-dimensionaler linearer Teilraum mit $0 \le h < k < n$.

Beispiel 12.8 $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ Design-Matrix
$Y = X\beta + \sigma W$
$L_k = \{X\beta : \beta \in \mathbb{R}^2\}, \quad L_h = \{X\tilde{\beta} : \tilde{\beta} \in \mathbb{R} \times \{0\}\} = \text{span}((1, \ldots, 1)^\top), \quad k = 2,\ h = 1$, falls x_i

nicht alle identisch.
Dann bedeutet $\mu \in L_h$ (oder $\beta \in \mathbb{R} \times \{0\}$), dass $Y = (\beta_0, \ldots, \beta_0)^\top + \sigma W$
Die Verteilung von Y hängt also nicht von $(x_1, \ldots, x_n)^\top$ ab.

Im Folgenden betrachte das

Testproblem (12.7)

$H = \{\mu \in L_h\}$ gegen $K = \{\mu \in L_k \setminus L_h\}$
Zu $y \in \mathbb{R}^n$ definiere
$\hat{\mu}(y) = \text{Proj}(y, L_k)$ orthogonale Projektion von y in L_k
$\hat{\tilde{\mu}}(y) = \text{Proj}(y, L_h)$
$\tilde{\tilde{\mu}}(y) = \text{Proj}(y, L_{k-h})$
wobei $L_{k-h} = L_h^\perp \cap L_k$ (orthogonales Komplement von L_h in L_k).

Es gilt:
$L_k = L_h \oplus L_{k-h}, \quad L_h \perp L_{k-h}$
und $\hat{\mu}(y) = \hat{\tilde{\mu}}(y) + \tilde{\tilde{\mu}}(y)$ oder $\tilde{\tilde{\mu}}(y) = \hat{\mu}(y) - \hat{\tilde{\mu}}(y)$
Das Testproblem (12.7) kann wie folgt formuliert werden:

Testproblem (12.8)

$H = \{\mu \in L_k : \tilde{\tilde{\mu}}(\mu) = 0\}$ gegen $K = \{\mu \in L_k : \tilde{\tilde{\mu}}(\mu) \neq 0\}$
Da $y \mapsto \tilde{\tilde{\mu}}(y)$ lineare Abbildung ist und $\tilde{\tilde{\mu}}(y) = \tilde{\tilde{\mu}}(\hat{\mu}(y))$ gilt, ist $\tilde{\tilde{\mu}}(y)$ optimaler Schätzer für $\tilde{\tilde{\mu}}(\mu)$ gemäß Satz 12.6. Die Verteilung von $\tilde{\tilde{\mu}}(y)$ hängt noch von der unbekannten Varianz σ^2 ab, die optimal durch $\hat{\sigma}^2 = \frac{1}{n-k}||y - \hat{\mu}(y)||^2$ geschätzt werden kann. Daher ist folgende Teststatistik sinnvoll:

$$T^*(y) = \begin{cases} \frac{\frac{1}{k-h}||\hat{\mu}(y) - \hat{\tilde{\mu}}(y)||^2}{\frac{1}{n-k}||y - \hat{\mu}(y)||^2} & \text{für } ||y - \hat{\mu}(y)|| > 0 \\ \infty & \text{für } ||y - \hat{\mu}(y)|| = 0,\ ||\hat{\mu}(y) - \hat{\tilde{\mu}}(y)|| > 0 \\ -1 & \text{für } ||y - \hat{\mu}(y)|| = 0,\ ||\hat{\mu}(y) - \hat{\tilde{\mu}}(y)|| = 0 \end{cases}$$

Verwerfe H, wenn T^* groß ist.
Beachte: $y - \hat{\tilde{\mu}}(y) = \underbrace{y - \hat{\mu}(y)}_{\in L_k^\perp} + \underbrace{\hat{\mu}(y) - \hat{\tilde{\mu}}(y)}_{\in L_{k-h}}$
Also: $||y - \hat{\tilde{\mu}}(y)||^2 = ||y - \hat{\mu}(y)||^2 + ||\hat{\mu}(y) - \hat{\tilde{\mu}}(y)||^2$

Bemerkung 12.9 $||y - \hat{\tilde{\mu}}(y)||^2$ (bzw. $||y - \hat{\mu}(y)||^2$) entspricht bis auf Vorfaktoren dem optimalen Schätzer für σ^2 unter H (bzw. $H + K$).

Definition 12.10 *Seien $Y_1, \ldots, Y_{r+s}$ stochastisch unabhängige Zufallsvariablen mit $\mathcal{L}(Y_i) = N_{(0,1)},\ 1 \leq i \leq r, \quad \mathcal{L}(Y_i) = N_{(\delta_i, 1)},\ r+1 \leq i \leq r+s.$*
Die Verteilung von

$$\frac{\frac{1}{s}\sum_{i=r+1}^{r+s} Y_i^2}{\frac{1}{r}\sum_{i=1}^{r} Y_i^2}$$

heißt ***nichtzentrale*** $F_{s,r}(\delta^2)$***-Verteilung*** *mit dem Nichtzentralitätsparameter*

$$\delta^2 = \sum_{i=1}^{s} \delta_{r+i}^2$$

Ist $\delta^2 = 0$*, so liegt eine zentrale* $F_{s,r}$*-Verteilung vor (vergleiche Definition 4.1 3.).*

Lemma 12.11 *Die Verteilung* $F_{s,r}(\delta^2)$ *hängt nur von* $\delta^2 = \sum_{i=1}^{s} \delta_{r+i}^2$ *ab (und nicht von* δ_i*).*

Beweis: Sei $\delta^2 > 0$ und $e_1, \ldots, e_s \in \mathbb{R}^s$ eine orthonormale Basis des $\mathbb{R}^s$ mit $e_1 = \frac{1}{\delta}(\delta_{r+1}, \ldots, \delta_{r+s})^\top$.
Sei $\mathcal{T} = (e_1, \ldots, e_n)^\top$ orthogonale Matrix und $Z = (Z_1, \ldots, Z_s)^\top = (Y_{r+1} - \delta_{r+1}, \ldots, Y_{r+s} - \delta_{r+s})^\top$.
Dann $Z_1, \ldots, Z_s$ i.i.d. $N_{(0,1)}$-verteilt und $\mathcal{L}(\mathcal{T}Z) = \mathcal{L}(Z)$ (Satz 4.3).
Mit der Notation $d = (\delta_{r+1}, \ldots, \delta_{r+s})^\top$ gilt:
$\sum_{i=r+1}^{r+s} Y_i^2 = ||Z + d||^2 = (Z+d)^\top \underbrace{\mathcal{T}^\top \mathcal{T}}_{I_s}(Z+d) = ||\mathcal{T}(Z+d)||^2 = ||Z + (\delta, 0, \ldots, 0)^\top||^2 =$
$(Z_1 + \delta)^2 + \sum_{i=2}^{s} Z_i^2$ □

Satz 12.12 *Es liege ein lineares Modell mit Normalverteilungsannahme vor.*

1. *Unter* (μ, σ), $\sigma \neq 0$, *besitzt* $T^*(Y)$ *eine nichtzentrale* $F_{k-h,n-k}(\delta^2)$*-Verteilung mit* $\delta^2 = \frac{||\tilde{\tilde{\mu}}(\mu)||^2}{\sigma^2}$.

2. *Es bezeichne* c_α *das* α*-Fraktil der zentralen* $F_{k-h,n-k}$*-Verteilung. Dann bildet*

$$\varphi^*(y) = \begin{cases} 1 & > \\ T^*(y) & c_\alpha \\ 0 & \leq \end{cases}$$

(F-Test der linearen Hypothese) einen Test für $H = L_h$ *gegen* $K = L_k \setminus L_h$ *zum Niveau* α.

Beweis:

1. Wir gehen zum kanonischen Modell über: $\tilde{Y} = \mathcal{T}Y = \mathcal{T}\mu + \sigma\mathcal{T}W = \nu + \sigma\tilde{W}$, wobei $e_1, \ldots, e_n$ eine Orthonormalbasis sei mit $L_h = \text{span}(e_1, \ldots, e_h)$ und $L_k = \text{span}(e_1, \ldots, e_k)$, also $\mathcal{T}L_h = \mathbb{R}^h \times \{0\}^{n-h}$, $\mathcal{T}L_k = \mathbb{R}^k \times \{0\}^{n-k}$
   $\hat{\tilde{\nu}}(\tilde{y}) = (\tilde{y}_1, \ldots, \tilde{y}_h, 0, \ldots, 0)^\top$, $\quad \hat{\nu}(\tilde{y}) = (\tilde{y}_1, \ldots, \tilde{y}_k, 0, \ldots, 0)^\top$, $\quad \tilde{y} \in \mathbb{R}^n$
   Wegen $\hat{\mu} = \mathcal{T}\hat{\nu}$ und $\hat{\tilde{\mu}} = \mathcal{T}\hat{\tilde{\nu}}$ gilt:
   $T^*(y) = \frac{\frac{1}{k-h}||\hat{\mu}(y) - \hat{\tilde{\mu}}(y)||^2}{\frac{1}{n-k}||y - \hat{\mu}(y)||^2} = \frac{\frac{1}{k-h}||\hat{\nu}(\tilde{y}) - \hat{\tilde{\nu}}(\tilde{y})||^2}{\frac{1}{n-k}||\tilde{y} - \hat{\nu}(\tilde{y})||^2} = \frac{\frac{1}{k-h}\sum_{i=h+1}^{k} \tilde{y}_i^2 \cdot \frac{1}{\sigma^2}}{\frac{1}{n-k}\sum_{i=k+1}^{n} \tilde{y}_i^2 \cdot \frac{1}{\sigma^2}}$
   Da $\mathcal{T}$ orthogonal ist, sind $\tilde{Y}_1, \ldots, \tilde{Y}_n$ stochastisch unabhängig (Satz 4.3) mit
   $\mathcal{L}(\tilde{Y}_i) = N_{(\nu_i, \sigma^2)}$, $\quad \nu = \mathcal{T}\mu = (\nu_1, \ldots, \nu_n)^\top \in \mathbb{R}^k \times \{0\}^{n-k}$, also $\nu_{k+1}, \ldots, \nu_n = 0$.
   Dann besitzt $T^*(\tilde{Y})$ eine nichtzentrale $F_{k-h,n-k}(\delta^2)$-Verteilung mit
   $\delta^2 = \sum_{i=h+1}^{k} \frac{\nu_i^2}{\sigma^2} = \frac{||\nu - \hat{\tilde{\nu}}(y)||^2}{\sigma^2} = \frac{||\tilde{\tilde{\mu}}(\mu)||^2}{\sigma^2}$
2. Für $\mu \in \mathcal{L}_h$ gilt:
   $\nu_{h+1}, \ldots, \nu_n = 0 \Rightarrow \delta^2 = 0$ □

Beispiel 12.13 Wir betrachten ein Zweistichprobenproblem unter einer Normalverteilungsannahme (vergleiche Beispiel 12.7).
$Y'_{ji} = \mu_j + \sigma W'_{ji} \quad j = 1,2,\ 1 \le i \le n_j,\ n = n_1 + n_2$
Setze $Y_i = \begin{cases} Y'_{1i} & 1 \le i \le n_1 \\ \text{für} & \\ Y'_{2,i-n_1} & n_1 < i \le n \end{cases}$, $W_i = \begin{cases} W'_{1i} & 1 \le i \le n_1 \\ \text{für} & \\ W'_{2i} & n_1 < i \le n \end{cases}$,
also $Y = (\mu_1, \ldots, \mu_1, \mu_2, \ldots, \mu_2)^\top + \sigma W$

Testproblem: $\{\mu_1 = \mu_2\}$ gegen $\{\mu_1 \neq \mu_2\}$, $k = 2,\ h = 1$
$\mathcal{L}_k = \text{Span}((\underbrace{1, \ldots, 1}_{n_1}, 0, \ldots, 0)^\top, (0, \ldots, 0, \underbrace{1, \ldots, 1}_{n_2})^\top)$
$\mathcal{L}_h = \text{Span}((1, \ldots, 1)^\top) \subset \mathcal{L}_k$
Mit (12.7) folgt:
$\hat{\mu}(Y) = (\overline{Y}'_{1\cdot}, \ldots, \overline{Y}'_{1\cdot}, \overline{Y}'_{2\cdot}, \ldots, \overline{Y}'_{2\cdot})$ mit $\overline{Y}'_{j\cdot} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y'_{ji}$ und
$\hat{\sigma}^2(Y) = \frac{1}{n-2} \left(\sum_{i=1}^{n_1} (Y'_{1i} - \overline{Y}'_{1\cdot})^2 + \sum_{i=1}^{n_2} (Y'_{2i} - \overline{Y}'_{2\cdot})^2 \right)$
Zur Bestimmung von $\hat{\hat{\mu}}$ verwenden wir (12.7), da für $\mu_1 = \hat{\mu}_2$ ein Einstichprobenproblem vorliegt.
Es gilt $\hat{\hat{\mu}}(Y) = (\overline{Y}, \ldots, \overline{Y})$
Die Teststatistik besitzt dann die Form
$T^*(Y) = \frac{n_1 (\overline{Y} - \overline{Y}'_{1\cdot})^2 + n_2 (\overline{Y} - \overline{Y}'_{2\cdot})^2}{\hat{\sigma}^2(Y)} = \frac{\frac{n_1 n_2}{2} (\overline{Y}'_{1\cdot} - \overline{Y}'_{2\cdot})^2}{\hat{\sigma}^2(Y)}$
Damit erhalten wir den zweiseitigen F-Test für das Zweistichprobenproblem als Analogon zum einseitigen Zweistichproben-t-Test bei Normalverteilungen (vergleiche Abschnitt 4)).

Beispiel 12.14 (Einfache lineare Regression) Gegeben: $Y_i = \beta_0 + \beta_1 x_i + \sigma W_i, \quad W_1, \ldots, W_n$ i.i.d. $N_{(0,1)}$-verteilt.
Testproblem: $H = \{\beta_1 = 0\}$ gegen $K = \{\beta_1 \neq 0\}$
Setze $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ Design-Matrix
$Y = X \binom{\beta_0}{\beta_1} + \sigma W$
$L_2 = \text{Span}((1, \ldots, 1)^\top, (x_1, \ldots, x_n)^\top)$
$L_1 = \text{Span}((1, \ldots, 1)^\top)$
$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^\top$ mit $\hat{\beta}_0 = \overline{Y} - \overline{x}\ \hat{\beta}_1,\ \hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \overline{Y}) x_i}{\sum_{i=1}^n (x_i - \overline{x})^2}$
$\hat{\mu}(Y) = X\hat{\beta} = (\overline{Y} + \hat{\beta}_1 (x_1 - \overline{x}), \ldots, \overline{Y} + \hat{\beta}_1 (x_n - \overline{x}))$
Ferner $\hat{\hat{\mu}}(Y) = (\overline{Y}, \ldots, \overline{Y})^\top$
Also: $||\hat{\hat{\mu}}(Y) - \hat{\mu}(Y)||^2 = \hat{\beta}_1^2 - \sum_{i=1}^n (Y_i - \overline{x})^2$
$||\hat{\mu}(Y) - Y||^2 = \sum_{i=1}^n (Y_i - \overline{Y} - \hat{\beta}_1 (x_i - \overline{x}))^2$
$\Rightarrow T^*(Y) = \frac{\hat{\beta}_1 \sum_{i=1}^n (x_i - \overline{x})^2}{\sum_{i=1}^n \left((Y_i - \overline{Y}) - \hat{\beta}_1 (x_i - \overline{x}) \right)^2}$

Kapitel IV

Elemente der asymptotischen Statistik

13 Asymptotische Normalität von Maximum-Likelihood-Schätzern

Im Folgenden: $\mathcal{Q} = \{Q_\vartheta^n : \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen auf $(S^n, \mathcal{B}^n)$
μ σ-endliches dominierendes Maß von Q_ϑ $\forall \vartheta$
$f_\vartheta \in \frac{dQ_\vartheta}{d\mu}$ μ-Dichte von Q_ϑ
Dann $(x \mapsto \prod_{i=1}^n f_\vartheta(x_i)) \in \frac{dQ_\vartheta^n}{d\mu^n}$ Produktdichte.

Definition 13.1 *$\hat{\vartheta}_n : S^n \mapsto \Theta$ heißt **Maximum-Likelihood-Schätzer**[1] (zum Stichprobenumfang n), falls*

$$\prod_{i=1}^n f_{\hat{\vartheta}_n(x)}(x_i) = \sup_{\vartheta \in \Theta} \prod_{i=1}^n f_\vartheta(x_i)$$

Beispiel 13.2 Sei $n = 1$, $\Theta = \{0, 1\}$,
$Q_0(\{-1\}) = \frac{1}{3}$, $Q_0(\{1\}) = \frac{2}{3}$, $Q_1(\{-1\}) = \frac{3}{4}$, $Q_1(\{1\}) = \frac{1}{4}$
$\mu = \varepsilon_{-1} + \varepsilon_1$ Zählmaß.
Dann $f_i(x) = Q_i(\{x\})$
Falls -1 beobachtet wird, schätze $\hat{\vartheta}(-1) \in \{0, 1\}$ mit $f_{\hat{\vartheta}(-1)}(-1) = \sup\{f_0(-1), f_1(-1)\} = \frac{3}{4}$,
also $\hat{\vartheta}(-1) = 1$.

Bezeichnung 13.3 *Es heißen*

1. $\vartheta \mapsto \prod_{i=1}^n f_\vartheta(x_i)$ ***Likelihood-Funktion.***
2. $\vartheta \mapsto \sum_{i=1}^n \log f_\vartheta(x_i)$ ***Log-Likelihood-Funktion.***
3. $\sum_{i=1}^n \frac{\partial \log}{\partial \vartheta} f_\vartheta(x_i)$ ***Likelihood-Gleichung.***

[1] im Weiteren als *ML-Schätzer* bezeichnet

Beispiel 13.4 (Poisson-Verteilungen mit Parameter $\lambda > 0$) $(S, \mathcal{B}) = (\mathbb{N}_0, \mathbb{P}(\mathbb{N}_0))$, μ Zählmaß auf $\mathbb{N}_0$
$P_\lambda(\{x\}) = f_\lambda(x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x \in \mathbb{N}_0$
Dann $\prod_{i=1}^n f_\lambda(x_i) = \max_\lambda! \Leftrightarrow \sum_{i=1}^n (x_i \log(\lambda) - \lambda - \log(x_i!)) = \max_\lambda!$
Diff. $\Rightarrow \sum_{i=1}^n \frac{\partial f_\lambda(x_i)}{\partial \lambda} = \sum_{i=1}^n (\frac{x_i}{\lambda} - 1) \stackrel{!}{=} 0 \Leftrightarrow \sum_{i=1}^n x_i = n\lambda$
$\Leftrightarrow \lambda_n = \frac{1}{n}\sum_{i=1}^n x_i$ ML-Schätzung.

Wegen $\int x\ dP_\lambda(x) = \lambda$ und $\int (x-\lambda)^2 dP_\lambda(x) = \lambda$ gilt nach dem Zentralen Grenzwertsatz

$$P_\lambda^n \left\{ x \in S^n : \frac{\sqrt{n}}{\sqrt{\lambda}}(\hat{\lambda}_n(x) - \lambda) \le t \right\} \stackrel{n\to\infty}{\longrightarrow} \Phi(t), \quad t \in \mathbb{R}$$

„asymptotische Normalität“

Daraus folgt:
$P_\lambda^n\{x \in \mathbb{N}_0^n : |\hat{\lambda}_n - \lambda| \le \frac{t\sqrt{\lambda}}{\sqrt{n}}\} \stackrel{n\to\infty}{\longrightarrow} \Phi(t) - \Phi(-t) = 2\Phi(t) - 1$ und somit

$$\forall \varepsilon > 0 \qquad P_\lambda^n\{x \in \mathbb{N}_0^n : |\hat{\lambda}_n - \lambda| \le \varepsilon\} \longrightarrow 1$$

„schwache Konsistenz“

Satz 13.5 (Asymptotische Normalität des ML-Schätzers) *Es sei $\Theta \subset \mathbb{R}$ offen, $\vartheta_0 \in \Theta$.*

1. *Sei $U = U(\vartheta_0)$ eine Umgebung von ϑ_0, so dass $\vartheta \mapsto f_\vartheta(x)$ auf U dreifach stetig differenzierbar ist für alle $x \in S$.*
2. *Es gebe eine Q_{ϑ_0}-integrierbare Funktion φ_{ϑ_0} mit*
   $\forall x \in S: \quad \sup_{\vartheta \in U} |\frac{\partial^3}{\partial \vartheta^3} \log f_\vartheta(x)| \le \varphi_{\vartheta_0}(x)$.
3. $\int \frac{\frac{\partial^2}{\partial \vartheta^2} f_{\vartheta_0}}{f_{\vartheta_0}} dQ_{\vartheta_0} = 0, \qquad \int \frac{\frac{\partial}{\partial \vartheta} f_{\vartheta_0}}{f_{\vartheta_0}} dQ_{\vartheta_0} = 0$
4. *Für die Fisher-Information $I(\vartheta) = \int (\frac{\partial}{\partial \vartheta} \log f_\vartheta)^2 dQ_\vartheta$ gelte: $0 < I(\vartheta_0) < \infty$*

Dann gilt für jede schwachkonsistente Folge $\hat{\nu}_n,\ n \in \mathbb{N}$, von ML-Schätzern:

$$\forall t \in \mathbb{R}: \qquad Q_{\vartheta_0}^n\{x \in S^n : n^{1/2} I(\vartheta_0)^{1/2}(\hat{\nu}_n(x) - \vartheta_0) \le t\} \longrightarrow \Phi(t)$$

Beweis: Ein ML-Schätzer ist eine Lösung der ML-Gleichung
$\sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_\vartheta(x_i) = 0$
Damit gilt mittels einer Taylor-Entwicklung um ϑ_0:
$0 = \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_{\hat{\vartheta}_n(x)}(x_i) = \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_{\vartheta_0}(x_i) + \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_{\tilde{\vartheta}_n(x)}(x)(\hat{\vartheta}_n(x) - \vartheta_0)$
wobei $\tilde{\vartheta}_n(x)$ zwischen $\hat{\vartheta}_n$ und ϑ_0 liegt.
Damit $\hat{\vartheta}_n(x) - \vartheta_0 = \frac{n^{-1/2}\sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_{\vartheta_0}(x_i)}{\frac{1}{n} - \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f_{\tilde{\vartheta}_n(x)}(x_i)}$
Wir zeigen:

1. $\int(\frac{\partial}{\partial\vartheta}\log f_{\vartheta_0})dQ_{\vartheta_0}=0$
2. $Q^n_{\vartheta_0}\{x\in S^n: I(\vartheta_0)^{-1/2}n^{-1/2}\sum_{i=1}^n\frac{\partial}{\partial\vartheta}\log f_{\vartheta_0}(x_i)\le t\}\overset{n\to\infty}{\longrightarrow}\Phi(t),\quad t\in\mathbb{R}$
3. $\forall\varepsilon>0:\qquad Q^n_{\vartheta_0}\{x\in S^n:|-\frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\vartheta}\log f_{\tilde\vartheta_n(x)}(x_i)-I(\vartheta_0)|\le\varepsilon\}\longrightarrow 1$

Dann folgt die Behauptung aus dem Lemma von Slutsky (siehe Übung).

zu 1: Folgt aus Voraussetzung 3 mit $\frac{\partial}{\partial\vartheta}\log f_\vartheta|_{\vartheta=\vartheta_0}=\frac{\frac{\partial}{\partial\vartheta}f_{\vartheta_0}}{f_{\vartheta_0}}|_{\vartheta=\vartheta_0}$[2]

zu 2: Folgt unmittelbar aus dem Zentralen Grenzwertsatz:
$Z_i(x)=\frac{\partial}{\partial\vartheta}\log f_{\vartheta_0}(x_i),\quad i=1,\dots,n$, unter $Q^n_{\vartheta_0}$ unabhängig mit
$E_{Q^n_{\vartheta_0}}Z_i=0,\quad Var_{Q^n_{\vartheta_0}}(Z_i)=\int(\frac{\partial}{\partial\vartheta}\log f_{\vartheta_0})^2dQ_{\vartheta_0}=I(\vartheta_0)$

zu 3: Nach dem schwachen Gesetz der großen Zahl:
$\forall\varepsilon>0:\qquad Q_{\vartheta_0}\{x\in S^n:|-\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\vartheta^2}\log f_{\vartheta_0}(x_i)-\mu_0|>\varepsilon\}\longrightarrow 0$

wobei $\mu_0=-\int(\frac{\partial^2}{\partial\vartheta^2}\log f_{\vartheta_0})dQ_{\vartheta_0}=-\int\left(\frac{\partial}{\partial\vartheta}\frac{\frac{\partial}{\partial\vartheta}f_\vartheta}{f_\vartheta}\right)|_{\vartheta=\vartheta_0}dQ_{\vartheta_0}=-\int\left(\frac{\frac{\partial^2}{\partial\vartheta^2}f_\vartheta}{f_\vartheta}-\frac{(\frac{\partial}{\partial\vartheta}f_\vartheta)^2}{f^2_\vartheta}\right)|_{\vartheta=\vartheta_0}dQ_{\vartheta_0}\overset{Vor.\ 3}{=}$

$\int\left(\frac{(\frac{\partial}{\partial\vartheta}f_\vartheta)^2}{f^2_\vartheta}|_{\vartheta=\vartheta_0}\right)dQ_{\vartheta_0}=\int(\frac{\partial}{\partial\vartheta}\log f_\vartheta|_{\vartheta=\vartheta_0})^2dQ_{\vartheta_0}=I(\vartheta_0)$

Aus dem Mittelwertsatz der Differentialrechnung folgt mit Voraussetzung 2:
$|\frac{\partial^2}{\partial\vartheta^2}\log f_{\hat\vartheta_n(x)}(x_i)-\frac{\partial^2}{\partial\vartheta^2}\log f_{\vartheta_0}(x_i)|\le\varphi_{\vartheta_0}(x_i)|\tilde\vartheta_n(x)-\vartheta_0|\le\varphi_{\vartheta_0}(x_i)|\hat\vartheta_n(x)-\vartheta_0|$
Aus dem schwachen Gesetz der großen Zahl, dem Lemma von Slutsky und der schwachen Konsistenz von $\hat\vartheta_n$ folgt
$\forall\varepsilon>0:\qquad Q^n_{\vartheta_0}\{x\in S^n:|\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\vartheta^2}\log f_{\tilde\vartheta_n(x)}(x_i)-\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\vartheta^2}\log f_{\vartheta_0}(x_i)|>\varepsilon\}\le Q^n_{\vartheta_0}\{x\in S^n:|\underbrace{\frac{1}{n}\sum_{i=1}^n\varphi_{\vartheta_0}(x_i)}_{\to E_{Q_{\vartheta_0}}\varphi_{\vartheta_0}}||\underbrace{\hat\vartheta_n(x)-\vartheta_0}_{\to 0}|>\varepsilon\}\longrightarrow 0$

Damit folgt 3. □

[2] $\frac{\partial}{\partial\vartheta}f_{\vartheta_0}=\frac{\partial}{\partial\vartheta}f_\vartheta|_{\vartheta=\vartheta_0}$

Anhang A

Ergänzungen aus den Übungen

Beweis zu Korollar 2.6

1. $\psi = \varphi$ ist Test z.N. α, also $E_1\varphi \geq E_1\psi = \alpha$
2. Sei φ_α NP-Test aus Lemma 2.4. Dann gilt nach Lemma 2.4 1. $0 = E_1\varphi - E_1\varphi_\alpha \leq c(E_0\varphi - \underbrace{E_0\varphi_\alpha}_{\alpha})$, also $c = 0 \vee E_0\varphi = \alpha$. Wegen $c = 0 \Rightarrow E_0\varphi_\alpha = 1$ folgt $E_0\varphi = \alpha$.

   a) $E_1\varphi > \alpha$. Dann gilt wegen $E_0\varphi = \alpha \neq E_1\varphi \qquad Q_0 \neq Q_1$

   b) $E_1\varphi \leq \alpha$, also wegen 1. $E_1\varphi = \alpha$. Sei $\psi \equiv \alpha$. Dann gilt $0 \geq (\psi - \varphi)(f_1 - cf_0) \Rightarrow 0 \geq E(\psi - \varphi)(f_1 - cf_0) = \underbrace{E_1\psi - E_1\varphi}_{0} - c\underbrace{(E_0\psi - E_0\varphi)}_{0} = 0$, also $(\psi - \varphi)(f_1 - cf_0) = 0 \quad \mu$-f.ü.

   $\Rightarrow f_1 = cf_0 \quad \mu$-f.ü., also $c = 1$ und $Q_1 = Q_0$. □

Beweis zu Bemerkung 3.2

(i) $Q_1 \ll Q_0$, d.h. $Q_0(A) = 0 \Rightarrow Q_1(A) = 0$, und $\exists$ Dichte f so dass $Q_1(A) = \int\limits_A f \, dQ_0$

$Q_1(\{L = \infty\}) = \int\limits_{\{L=\infty\}} L(x)dQ_0(x) + Q_1(\{L = \infty\}) \Leftrightarrow \int\limits_{\{L=\infty\}} L(x)Q_0(dx) = 0$

$\Rightarrow L = f \quad Q_0$-f.ü.,und L Q_0-Dichte von Q_1.

(ii) $\int\limits_A \frac{q_1(x)}{q_0(x)} 1_{(0,\infty)}(q_0(x))dQ_0(x) + Q_1(A \cap \{L = \infty\}) = \int\limits_A q_1(x)1_{(0,\infty)}(q_0(x))d\mu(x) + Q_1(A \cap \{q_0 = 0\}) = \int\limits_A 1_{(0,\infty)}(q_0(x))dQ_1(x) + \int\limits_A 1_{\{0\}}(q_0(x))dQ_1(x) = \int\limits_A Q_1(dx) = Q_1(A)$ □

Beweis zu Korollar 7.5

(i) Nach Satz 7.4 gilt: $Q(Z \in \cdot | T = t)$ existiert unabhängig von ϑ mit der Eigenschaft
$(Z,T)Q_\vartheta(A \times B) = Q(Z \in A, T \in B) = \int_B Q_\cdot(Z \in A|T = t)TQ_\vartheta(dt)$
Da $f(Z,T)$ integrierbar bzgl. Q_ϑ, folgt mit Fubini für Markovkerne $\infty > \int f(Z,T)dQ_\vartheta = \int f(z,t)d(Z,T)Q_\vartheta(z,t) \overset{Fubini}{=} \iint f(z,t)Q_\cdot(Z \in dz|T = t)TQ_\vartheta(dt)$

(ii) Mit Lemma 6.5 folgt $E(f(Z,T)|T = t) = \int f(z,t)Q_\cdot(Z \in dz|T = t)$ □

Lemma 1.1 *Seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen und $h(X_1, \ldots, X_n)$ integrierbar. Dann gilt:*

$$E(h(X_1, \ldots, X_n)|X_{1:n}, \ldots, X_{n:n}) = \frac{1}{n!} \sum_{\pi \in \Pi} h(X_{\pi(1)}, \ldots, X_{\pi(n)}),$$

wobei Π die Menge der Permutationen auf $\{1, \ldots, n\}$ bezeichnet.
Speziell gilt $E(X_1|X_{1:n}, \ldots, X_{n:n}) = \overline{X}$.

Beweis: Sei $Y := \frac{1}{n!} \sum_{\pi \in \Pi} h(X_{\pi(1)}, \ldots, X_{\pi(n)})$. Es ist $\pi : (\mathbb{R}^n, \mathbb{B}^n) \mapsto (\mathbb{R}^n, \mathbb{B}^n)$ mit $\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$, und es gilt
$(\pi \circ (X_1, \ldots, X_n))P = (X_1, \ldots, X_n)P$ (Permutationsinvarianz von Produktmaßen) $(*)$
Sei $Q := XP$, dann $(X_1, \ldots, X_n)P = Q^n$, da X_i i.i.d., und $(X_1, \ldots, X_n)P(B_1, \ldots, B_n) = \prod_{i=1}^n Q(B_i)$
Sei $T : (\mathbb{R}^n, \mathbb{B}^n) \mapsto (\mathbb{R}^n, \mathbb{B}^n)$ mit $T(x_1, \ldots, x_n) = (x_{1:n}, \ldots, x_{n:n})$. Dann $T \circ \pi = T \ \forall \pi \in \Pi$.

(i) Messbarkeit:
$Y = \frac{1}{n!} \sum_{\pi \in \Pi} h(X_{\pi(1)}, \ldots, X_{\pi(n)}) = \frac{1}{n!} \sum_{\pi \in \Pi} h \circ \pi \circ (X_{1:n}, \ldots, X_{n:n})$

(ii) $E_P(1_B(X_{1:n}, \ldots, X_{n:n}) \cdot h(X_1, \ldots, X_n)) = E_P(1_B \circ T \circ (X_1, \ldots, X_n) \cdot h(X_1, \ldots, X_n))$
$= E_P(1_B \circ (T \circ \pi) \circ (\pi^{-1}(X_1, \ldots, X_n)) \cdot (h \circ \pi) \circ (\pi^{-1}(X_1, \ldots, X_n))) \overset{(*)}{=} E_P(1_B \circ (T \circ \pi)(X_1, \ldots, X_n) \circ (h \circ \pi)(X_1, \ldots, X_n)) = \frac{1}{n!} \sum_{\pi \in \Pi} E_P(1_B \circ T(X_1, \ldots, X_n) \cdot (h \circ \pi)(X_1, \ldots, X_n)) = E_P(1_B(X_{1:n}, \ldots, X_{n:n}) \cdot \frac{1}{n!} \sum_{\pi \in \Pi} h(X_{\pi(1)}, \ldots, X_{\pi(n)}))$

(iii) Mit $h(X_1, \ldots, X_n) := X_1$ folgt
$E(X_1|X_{1:n}, \ldots, X_{n:n}) = \frac{1}{n!} \sum_{\pi \in \Pi} X_{\pi(1)} = \frac{1}{n!} (\sum_{i=1}^n X_i) \cdot (n-1)! = \overline{X}$ □

Beweis zu Beispiel 7.14

Mit obigem Lemma 1.1 gilt
$E_{Q^n}(1_B|T) = E_P(1_B(X_1, \ldots, X_n)|(X_{1:n}, \ldots, X_{n:n})) = \frac{1}{n!} \sum_{\pi \in \Pi} 1_B(X_{\pi(1)}, \ldots, X_{\pi(n)})$
$= E_\cdot(1_B|T)$ □

Beweis zu Lemma 11.1

Lemma 1.2 (Slutsky)

Anhang B

Literatur

1 Zur Statistik

[1] Falk, Becker, Marohn: *Angewandte Statistik mit SAS*, Springer 1995

[2] Pfanzagl, J.: *Allgemeine Methodenlehre der Statistik*, de Gruyter 1978

[3] Pruscha: *Angewandte Methoden der mathematischen Statistik*, Teubner 1989

[4] Witting: *Mathematische Statistik I*, Springer 1985

2 Zur Asymptotischen Statistik

[5] Pfanzagl: *Parametric Statistical Theorie*, de Gruyter 1994

[6] Rüschendorf: *Asymptotische Statistik*, 1988

[7] Witting: *Mathematische Statistik II*, 1995

Index

Äquivalenz, 27

Alternative, 5
asymptotische Normalität, 51

bedingte
    Erwartung, 23
    Verteilung, 23
    Wahrscheinlichkeit, 23
Bereich
    bester kritischer, 7
    gleichmäßig bester Konfidenz-, 20
    Konfidenz-, 20
    kritischer, 5
bias, 30

Design-Matrix, 38
    zufällige, 41
Dichtequotient, 10
    streng monotoner, 10
Differenzierbarkeit im quadratischen Mittel, 33

Exponentialfamilie, 12

Fehler
    1. Art, 5
    2. Art, 6
Fisher-Information, 33
Fraktil, 7
Funktion
    Güte-, 5, 6
    inverse Verteilungs-, 7
    kritische, 6
    Neyman-Pearson-, 7
    Quantil-, 7
    Risiko-, 29
    Verlust-, 29

Gegenhypothese, 5

Hypothese, 5

kritischer
    Bereich, 5
    Wert, 10

Lemma
    Neyman-Pearson-, 7
    von Slutsky, 54
Likelihood
    -Funktion, 50
    -Gleichung, 50
    Log-∼-Funktion, 50
    Maximum-∼-Schätzer, 50

Markov-Kern, 25

Neyman-Kriterium, 27

Parameterraum, 5

Quantil, 7

Regression
    -sgerade, 36
Residuen, 37
Risikofunktion, 29

Satz
    von Gauß-Markov, 40
    von Lehmann-Scheffé, 31
    von Rao-Blackwell, 30
Schätzer, 25
    erwartungstreuer, 40
    gleichmäßig bester minimum Varianz-, 31
    gleihmäßig bester erwartungstreuer, 31

Kleinste-Quadrate-, 37
linearer, 40
Maximum-Likelihood-, 50
Residual-, 44
schwache Konsistenz, 51
Statistik, 10, 25
Ordnungs-, 29
suffiziente, 26
statistisches
Experiment, 24
Funktional, 25
Stichprobenraum, 5
Suffizienz, 26

Test
F-, 18
F-$\sim$ der linearen Hypothese, 48
J-ähnlicher, 11
$\mathcal{X}^2$-, 16
t-, 16
bester, 6
Gauß-, 16
gleichmäßig bester, 7
nicht-randomisierter, 5
randomisierter, 6
unverfälschter, 11
zweiseitiger t-, 17
zweiseitiger Gauß-, 17
Zweistichproben-t-, 18
Zweistichproben-Gauß-, 17

Ungleichung
Chapman-Robbins-, 32
Cramér-Rao-, 34

Verlustfunktion, 29
Gaußsche, 29
konvexe, 30
Verteilung
$F_{s,r}$-
nichtzentrale, 48
zentrale, 15
$\mathcal{X}^2_n$-, 15
t_n-, 15
empirische bedingte, 43
Verteilungsfunktion
empirische bedingte, 43
Vollständigkeit, 30